Embodied Intelligent Robots: Technological Evolution, Core Systems, Challenges, and Future Outlook

Xin Yuan
Guangzhou Huali College, Guangzhou, Guangdong, China

Abstract: As an innovative product of multi-disciplinary integration, embodied intelligent robots break the functional limitations of traditional robots. By achieving in-depth interaction with the physical environment to realize autonomous decision-making, they have become a core support for intelligent transformation. This paper systematically sorts out their four major development stages, constructs the "perception-decision-action-feedback-learning" core technical system, and details key technologies such as multi-modal perception and "big brain-small brain" collaborative decision-making. Meanwhile, it analyzes four major current challenges: lack of standardization, dependence on core components, talent shortage, and ethical risks, and proposes five development directions: technological innovation, industrial collaboration, scene penetration, policy improvement, and human-robot symbiosis. The research provides multi-dimensional references for scientific research, industrial application, and policy formulation of embodied intelligent robots, helping them upgrade towards general intelligence and promoting the construction of a new production relationship of "technology empowerment and human leadership".

Keywords: Embodied Intelligent Robots; Embodied Intelligence; Disembodied Intelligence; General Intelligence; Human-Robot Symbiosis

1. Introduction

In the continuous development of the artificial intelligence field, embodied intelligent robots are gradually occupying a key position and becoming a core force driving industry progress. Essentially, embodied intelligent robots are a crystallization of multi-disciplinary technologies including artificial intelligence, mechanical engineering, and control science. They break the limitation of traditional robots that can only execute preset tasks, realize in-depth interaction with the physical environment and autonomous decision-making, and provide new solutions for the intelligent transformation of various industries.

This research aims to deeply analyze the technical system, development status, and future trends of embodied intelligent robots. Through systematic research, it comprehensively reveals the opportunities and challenges faced by embodied intelligent robots in technological innovation, application expansion, and industrial development, providing valuable reference for researchers, enterprise decision-makers, and policy-makers in related fields.

2. Origin and Development History of Embodied Intelligence

The concept of embodied intelligence can be traced back to the origin stage of artificial intelligence. In 1950, Turing proposed two development paths for intelligence in *Computing Machinery and Intelligence* [1], which later evolved into two paradigms: disembodied intelligence (focusing on algorithm optimization and symbolic reasoning) and embodied intelligence (emphasizing dynamic interaction between the physical body and the environment [2]). Its development process can be summarized into four core stages:

2.1 Theoretical Foundation and Early Practice

Gibson's ecological psychology [3] and Varela et al.'s "enactive cognition" [4] laid the theoretical foundation for embodied intelligence; Brooks' "subsumption architecture" [5] and Pfeifer et al.'s artificial life simulation [6] completed early technical practices.

2.2 Rise of Morphological Computation and Embodied Dynamics

The research focus shifted to "body-environment collaboration". The theory of morphological computation [2] advocates realizing intelligent behavior through physical structures. Studies by Collins et al. [7] and Pfeifer et al. [8] confirmed that using the interaction between morphology and the environment can simplify control logic and improve robot adaptability.

2.3 Establishment of Data-Driven Learning Paradigm

Deep learning and reinforcement learning have promoted the transformation of embodied intelligence towards "data-driven adaptability". Levine et al.'s deep visuomotor policies [9], Haarnoja et al.'s soft actor-critic algorithms [10], as well as projects such as OpenAI Dactyl [11] and Google RT-2 [12], have deepened the "perception-decision-action" closed-loop cognition and expanded the application boundaries.

2.4 Integration of Higher-Order Cognition and Causal Reasoning

Current research focuses on higher-order cognition and natural human-robot interaction. Schölkopf et al. [13] proposed causal representation learning, and Basu et al. [14] realized transfer learning with small-scale causal data. Today, embodied intelligence has gradually been applied in fields such as industrial automation and service robots, providing support for industrial innovation.

3. Key Technologies of Embodied Intelligent Robots

Embodied intelligent robots adopt the "perception-decision-action-feedback-learning" as the core closed-loop control logic. Through the collaborative linkage of each link, they realize autonomous operation in dynamic environments—they can not only actively capture multi-dimensional environmental information but also generate adaptive decisions based on preset algorithms, accurately execute tasks, and iteratively optimize in real time, ultimately achieving high flexibility and high reliability in operation. Their core technical system revolves around five key links:

3.1 Perception

Perception is the basic premise for the decision-making of embodied intelligent robots. It constructs environmental cognition through multi-source sensors and fusion technologies to ensure the rationality of subsequent decisions and actions.

3.1.1 Multi-modal information perception technology

Taking intelligent sensors with integrated data processing functions as the core carrier, matched with special sensors of different functions to meet the needs of complex scenes. Among them, MEMS sensors, with the advantages of small size, low cost, and low power consumption, can integrate multiple perception functions, adapting to the application of robots in narrow or dynamic scenes; brain-inspired vision chips such as "Tianmou Chip" developed by Yang [15] et al. can achieve high-speed, high-precision, and high dynamic range visual information collection under low bandwidth and low power consumption conditions; in addition, visual sensors undertake core tasks such as object recognition and posture tracking, force sensors support force control in precision machining, and tactile sensors can simulate biological skin to perceive material and grasping force. In the future, this technology will further integrate edge computing technology to improve real-time data processing capabilities.

3.1.2 Multi-modal information fusion technology

Comprehensively integrate different types of environmental information in stages to provide a comprehensive basis for accurate decision-making. According to the fusion stage, it can be divided into three types: early (raw data layer), middle (high-dimensional feature layer), and late (decision layer) [16]. Common methods include basic concatenation and weighted average, as well as cross-modal weight assignment, modal relationship graph construction, and multi-task collaborative optimization based on preset logic. It is worth noting that cross-modal alignment technology is a key prerequisite, which needs to establish semantic consistency of different modal data through preset rules to avoid information conflicts and ensure fusion effects.

3.2 Decision-Making

Breaking through the limitations of the traditional linear paradigm of "perception-planning-execution", it adopts a "big brain-small brain" collaborative architecture to improve generalization and response speed in dynamic scenes:

(1) Planning Layer (Brain)

As the core of decision-making, it undertakes high-level planning and overall behavior decision-making tasks, supported by preset logical algorithms and vision-mechanical collaborative models. During operation, it first integrates external environment description, visual input, spatial map, human instructions, and the robot's own state information; then, with the help of preset logical reasoning and task decomposition rules, it decomposes complex high-level goals into executable low-level action plans; at the same time, it dynamically adjusts the execution strategy by continuously updating environmental perception information to adapt to environmental changes and avoid the limitations of traditional fixed rules.

(2) Skill Layer (cerebellum)

Connecting decision-making planning and physical execution, it realizes the accurate implementation and optimization of skills through three types of traditional technologies:

Teaching and Reproduction Technology: The current mainstream implementation path, adopting the "teaching data collection + offline strategy solidification" model, including trajectory reproduction and multi-modal instruction mapping, providing a stable initial action plan for robots and laying the foundation for subsequent optimization.

Real-time Adjustment Technology: Through real-time interaction between the robot and the environment, dynamically optimize the action strategy based on preset feedback rules, focusing on skill adaptation and precision improvement, solving the problem of weak adaptability of traditional fixed strategies.

Traditional Control Technology: Embedded in the skill library in the form of "primitive skills" to ensure robustness in high-risk scenes. For example, fuzzy logic control adjusts clamping force through tactile feedback, model predictive control optimizes motion trajectory in real time, impedance control ensures the safety of human-robot collaboration, and industrial-grade PID control realizes stable positioning of structured tasks, providing deterministic support for overall action execution.

3.3 Action

As a key bridge between decision-making and physical execution, it undertakes the responsibility of converting high-level strategies into specific operations, directly affecting task execution efficiency and accuracy. Core technologies focus on path planning and motion control.

3.3.1 Path planning technology

The core goal is to plan the optimal route from the current position to the target position, which needs to take into account obstacle avoidance, dynamic environment adaptation, and robot kinematic constraints:

3.3.2 Environment modeling

Construct a map model that accurately reflects the real environment. Mainstream types include grid maps, topological maps, and feature maps.

3.3.3 Search algorithms

Traditional optimization algorithms are dominant to improve dynamic adaptability, such as A* algorithm (suitable for path search in static scenes), D* Lite algorithm (processing path updates in dynamic environments), and genetic algorithm, solving the problem of weak adaptability of traditional fixed paths.

3.3.4 Multi-robot collaboration

Divided into centralized and distributed types. Centralized planning achieves global optimization through a central control unit but has weak scalability; distributed coordination relies on information interaction between robots for autonomous coordination, with high scalability and robustness, suitable for large-scale multi-robot systems.

3.4 Feedback

As a key link connecting perception, decision-making, and action, it supports closed-loop control, dynamic environment adaptation, and system optimization, directly affecting task precision and robustness.

3.4.1 Model-Free adaptive control technology

Suitable for scenes where accurate modeling is difficult or the environment changes frequently. It does not require the establishment of an accurate mathematical model and realizes system adaptive adjustment through preset data processing rules. Core paths include action optimization based on trial-and-error correction, fuzzy control, and decision optimization based on statistical laws, greatly improving the robot's environmental adaptability and operational flexibility.

3.4.2 Model-based feedback control technology

Based on the premise of establishing accurate mathematical models of the robot and the environment, it ensures execution precision through two methods: one is impedance-admittance control, which dynamically adjusts control parameters to adapt to changes in environmental stiffness; the other is model predictive control, which carries out feedforward optimization combined with the system dynamics model, predicts the future trajectory first, and then adjusts in real time according to state differences to ensure efficient and accurate operation.

3.4.3 Virtual-real fusion dynamic adjustment technology

Integrate virtual and real data to optimize control strategies. Core paths include dynamic adjustment driven by multi-source data, AR and digital twin technology, and closed-loop feedback between simulation and physical operation, enabling stable operation of closed-loop control in complex environments.

3.5 Learning

Focusing on incremental optimization based on data accumulation, it is the key for robots to achieve long-term adaptation and performance improvement:

This technology aims to solve two major challenges of traditional systems: "difficulty in updating fixed strategies" and "slow adaptation to new scenes". Through real-time interaction between the robot and the environment, it continuously accumulates operational data and corrects action parameters, so that the optimization ability comes from the system's own operational experience rather than relying on external algorithm injection. Finally, the robot has continuous adaptability in dynamic environments, gradually optimizing the execution effect of new tasks and new scenes, providing support for long-term stable operation.

4. Current Challenges of Embodied Intelligent Robots

Core challenges focus on four aspects, directly affecting their large-scale application and high-quality development:

4.1 Lack of Standardization and Data Barriers

There is no unified technical standard or general development platform. Non-standard hardware interfaces and communication protocols lead to poor compatibility between devices from different manufacturers, making it difficult to collaborate; the high cost of dynamic interaction data collection, inconsistent data formats among enterprises form "data silos", and the lack of sharing mechanisms further exacerbates data scarcity, restricting the cross-scene generalization ability of technology.

4.2 Dependence on Core Components and Poor Software-Hardware Collaboration

Key components such as high-end graphics processing units and precision sensors have long been imported, which not only affects supply chain security but also limits independent industrial innovation; the decision-making system and motion control system adopt independent chip architectures, which are prone to communication delays. In complex working conditions such as industrial assembly and dynamic services, this will reduce the precision and stability of task execution.

4.3 Talent Shortage and Insufficient Scene Opening

Multi-disciplinary talents integrating AI, mechanical engineering, and control science are needed, but the education system is disconnected from industrial needs, resulting in lagging talent supply; the connection between R&D and application ends is not smooth, and public testing platforms are scarce. A large number of products remain in the laboratory stage, making it difficult to obtain real-scene data for iterative optimization.

4.4 Ambiguous Responsibility and Privacy Risks

There is a risk of leakage of private data collected during robot interaction, and the regulatory and protection system is not perfect; when accidents are caused by the autonomous decision-making of intelligent agents, there is no clear legal basis for the division of responsibilities among developers, operators, and users, leading to difficulties in accountability.

5. Future Outlook of Embodied Intelligent Robots

Breakthroughs around five directions to promote in-depth integration of technology and industry:

5.1 Technological Innovation Promotes Implementation

The coupling of multi-modal large models and world models promotes the upgrade of embodied intelligence from "single-task" to "general intelligence"; hardware breakthroughs such as bionic muscles and flexible sensors reduce manufacturing costs, laying the foundation for large-scale commercialization.

5.2 Industrial Ecosystem Collaboration

Policies promote standardization construction. Beijing, Zhejiang, and other places have piloted "unified hardware interfaces and standard data formats", and the first national-level development platform is expected to be launched in 2026; the industrial chain collaborates to tackle key problems, promoting robots to upgrade from "single-function" to "scene integration".

5.3 Full-Scene Application Penetration

In the industrial field, Tesla Optimus-like human-robot collaboration models are gradually promoted, and the global collaboration penetration rate in smart factories may exceed 40% by 2030; in the service field, robots enter inclusive scenes such as elderly care and education.

5.4 Improvement of Policies and Ethics

The *Regulations on the Safety Management of Embodied Intelligence* will be introduced to clarify the responsibility division and the "algorithm transparency + emergency fusing" mechanism; Beijing and Shanghai have piloted human-robot collaboration safety certification to balance technological innovation and social acceptance.

5.5 Human-Robot Symbiosis

With "human decision-making + machine execution" as the core, a new production relationship of "technology empowerment and human leadership" will be formed to achieve complementary win-win results between humans and machines.

6. Conclusion

As an innovative product integrating multi-disciplinary technologies, the development history of embodied intelligent robots has witnessed breakthroughs from theoretical germination to technical landing. The "perception-decision-action-feedback-learning" core technical system has built its ability to autonomously adapt to dynamic environments, while challenges such as lack of standardization and dependence on core components have clarified the key obstacles to industrial advancement. From technological innovation driving large-scale commercialization, to the industrial ecosystem moving towards collaboration, and then to full-scene application penetration and improvement of the policy and ethical system, embodied intelligent robots are moving towards the ultimate goal of "human-robot symbiosis". They not only provide intelligent solutions for fields such as industrial automation and people's livelihood services but also reshape the new production relationship of "technology empowerment and human leadership".

Through systematic analysis of the origin, technology, challenges, and outlook of embodied intelligent robots, this research provides multi-dimensional references for related fields. In the future, it is necessary to continuously focus on core technology research to break industrial bottlenecks, accelerate standardization construction to break collaboration barriers, and strengthen talent training and scene opening to promote in-depth alignment between technological innovation and social needs. It is believed that with the deepening of policy support and the collaborative upgrading of the industrial chain, embodied intelligent robots will occupy an important position in global technological competition and inject lasting impetus into the sustainable development of an intelligent society.

References

[1] TURING A M. Computing machinery and intelligence[J]. Mind, 1950, LIX(236): 433-460.

[2] LIU Huaping, GUO Di, SUN Fuchun, et al. Morphology-based embodied intelligence: Historical retrospect and research progress[J]. Acta Automatica Sinica, 2023, 49(6): 1131−1154.

[3] GIBSON J J. The ecological approach to visual perception: Classic edition[M]. Psychology Press, 1979.

[4] VARELA F J, THOMPSON E, ROSCH E. The embodied mind: Cognitive science and human experience[M]. The MIT Press, 1991.

[5] BROOKS R. A robust layered control system for a mobile robot[J]. IEEE Journal on Robotics and Automation, 1986, 2(1): 14-23.

[6] PFEIFER R. Understanding intelligence[M]. Cambridge, Massachusetts: MIT Press, 1999.

[7] COLLINS S, RUINA A, TEDRAKE R, et al.

Efficient bipedal robots based on passive-dynamic walkers[J]. Science, 2005, 307(5712): 1082-1085.

[8] PFEIFER R. How the body shapes the way we think: A new view of intelligence[M]. Cambridge: MIT Press, 2006.

[9] LEVINE S, FINN C, DARRELL T, et al. End-to-end training of deep visuomotor policies[J]. Journal of Machine Learning Research, 2016, 17(39): 1-40.

[10] HAARNOJA T, ZHOU A, HARTIKAINEN K, et al. Soft actor-critic algorithms and applications[EB/OL]. (2023-04-18). http://arxiv.org/abs/1812.05905.

[11] OpenAI, ANDRYCHOWICZ M, BAKER B, et al. Learning dexterous in-hand manipulation[EB/OL]. (2025-02-14). http://arxiv.org/abs/1808.00177.

[12] ZITKOVICH B, YU T, XU S, et al. RT-2: Vision-language-action models transfer web knowledge to robotic control[C]//Proceedings of The 7th Conference on Robot Learning. PMLR, 2023: 2165-2183.

[13] SCHÖLKOPF B, LOCATELLO F, BAUER S, et al. Towards causal representation learning[EB/OL]. (2025-02-14)[2025-02-14]. http://arxiv.org/abs/2102.11107.

[14] BASU S, TATLIDIL S, KIM M H, et al. Invited: Using causal information to enable more efficient robot operation[C]//Proceedings of the 61st ACM/IEEE Design Automation Conference. New York, NY, USA: Association for Computing Machinery, 2024: 1-4.

[15] YANG Z, WANG T, LIN Y, et al. A vision chip with complementary pathways for open-world sensing[J]. Nature, 2024, 629(8014): 1027-1033.

[16] WEI Shaomeng, ZHANG Jiangmin, SHI Hui, et al. Remaining useful life prediction of relative density high-dimensional kernel estimation for systems based on multi-sensor data fusion[J]. Computer Integrated Manufacturing System, 2024, 30(12): 4493-4507.

