Available Online: 8 September 2026 DOI: 10.54254/2977-3903/2026.36645

Review of key technologies for robot embodied intelligence oriented toward flexible manufacturing

Zheng Chen

School of International Education, Beijing University of Chemical Technology, Beijing, China

jackychen2004@qq.com

Abstract. Flexible manufacturing, characterized by high-mix, low-volume, and highly variable production, demands robotic systems with strong adaptability, dexterity, and intelligence that conventional offline-programmed industrial robots cannot provide. This paper presents a systematic review of key technologies for robot embodied intelligence oriented toward flexible manufacturing, organized around the closed loop of perception, decision-making, and execution. The purpose is to clarify the current research landscape, identify core technical bottlenecks, and outline future directions for embodied-intelligent manufacturing. Adopting a literature-analysis and comparative-review method, the study examines representative advances at three levels: multimodal environmental perception and real-time modeling, flexible adaptive precision manipulation, and intelligent decision-making for process planning and scheduling. The review finds that multimodal fusion and semantic Simultaneous Localization and Mapping (SLAM) are overcoming perception bottlenecks, that deep learning and force/position hybrid control are balancing flexible adaptability with high-precision operation, and that deep reinforcement learning and large models are advancing intelligent process planning. It concludes that data scarcity, model reliability, software-hardware integration, and ethical-legal standards remain the principal challenges to large-scale industrial deployment.

Keywords: embodied intelligence, flexible manufacturing, multimodal perception, robotic manipulation

1. Introduction

Driven by breakthrough advances in the new generation of artificial intelligence, embodied intelligence is rapidly permeating the industrial manufacturing sector. Embodied intelligence emphasizes that an intelligent agent interacts with the environment through a physical body, forming a closed loop of perception, learning, and decision-making during interaction—a concept that closely aligns with the demands of flexible manufacturing for robotic system flexibility, adaptability, and intelligence. The core of flexible manufacturing lies in coping with high-mix, low-volume, and highly variable production modes, requiring manufacturing systems to respond quickly to market changes and adjust process routes flexibly. Traditional industrial robots, which rely on offline programming and rigid control paradigms, cannot satisfy the requirements of flexible manufacturing for adaptive perception, dexterous manipulation, and intelligent decision-making.

The introduction of embodied intelligence offers new technological pathways for flexible manufacturing [1]. Its essence is to endow robotic systems with the closed-loop capability of "perception–decision–

execution," enabling them to interact with dynamic environments based on their own physical structures and to achieve adaptive learning, collaborative operation, and flexible response for complex tasks. Currently, embodied intelligence has demonstrated significant application value in manufacturing scenarios such as intelligent assembly, flexible processing, predictive maintenance, unmanned logistics, and intelligent inspection [2]. However, industrial embodied intelligence in flexible manufacturing environments still faces three major challenges: the difficulty of precise process modeling and monitoring under constrained perception, the challenge of balancing flexible adaptability with high-precision manipulation, and the difficulty of synergistically integrating general skills with specialized process knowledge [1].

From the perspective of closed-loop optimization of "perception–decision–execution," this paper systematically reviews the research progress of key technologies for robot embodied intelligence oriented toward flexible manufacturing, elaborates on the current state of perception, control, and decision-making layers, and analyzes future development trends. This work has both practical and academic significance: at the industrial level, systematically reviewing these key technologies helps manufacturing enterprises clarify the technological roadmap from environmental perception, flexible manipulation, to intelligent decision-making, providing technical references for intelligent transformation, cost reduction, efficiency improvement, and flexible upgrading of production lines under high-mix, low-volume production modes, thus contributing to the high-end and intelligent transformation of manufacturing. At the academic level, by structuring the review around the closed-loop optimization framework, this paper summarizes the research status, key bottlenecks, and trends at each level, offering systematic literature references and entry points for subsequent researchers in industrial embodied intelligence, robotic manipulation, and intelligent manufacturing, facilitating a quick grasp of the field landscape and identification of urgent research directions.

2. Robot multimodal environmental perception and real-time modeling

This section focuses on multimodal perception and real-time modeling of the flexible manufacturing environment by robots. Real-time modeling refers to the online construction and dynamic updating of geometric structures, poses, and semantic information models of the environment and workpieces during operation (e.g., via SLAM and 3D reconstruction techniques), thereby providing accurate and reliable environmental representations for subsequent flexible manipulation and intelligent decision-making.

2.1. Challenges for perception systems in flexible manufacturing environments

Flexible manufacturing environments are dynamically unstructured, characterized by: frequent changes in workpiece types leading to drastic variations in visual feature distributions; strong environmental interferences such as cutting fluid splashes, welding arcs, and smoke during processing; severe degradation of visual sensor imaging quality due to highly reflective metal surfaces; and mutual occlusion caused by multi-robot collaboration in confined workstations. These factors collectively constitute the challenge of precise process modeling and monitoring under constrained perception—the primary challenge for industrial embodied intelligence [1]. Traditional single-sensor perception schemes often fail in such complex environments, necessitating the development of multimodal, robust intelligent perception technologies.

2.2. Multisensor fusion perception architectures

In multimodal perception fusion, the collaboration of multiple sensors—vision, LiDAR, tactile, and force sensors—has become the mainstream approach. Vision sensors provide rich texture and color information, suitable for object recognition and pose estimation; LiDAR provides precise geometric structure measurements and remains stable under severe illumination changes; tactile and force sensors directly perceive physical quantities in contact interactions, providing feedback for precision operations. The major research program "Fundamental Theories and Key Technologies of Collaborative Robots" explicitly identifies "human-robot-environment multimodal perception and natural interaction" as one of its core scientific issues [3]. Typical multimodal fusion architectures include early fusion (data-level fusion), feature-level fusion, and late fusion (decision-level fusion). Early fusion preserves the integrity of raw information but incurs high computational costs; late fusion is efficient but may lose correlations. Researchers often choose or combine different fusion strategies according to specific task requirements.

2.3. Semantic SLAM technology in dynamic environments

Simultaneous Localization and Mapping (SLAM) is fundamental to autonomous robot navigation. In dynamic environments, traditional SLAM algorithms assume static scenes; moving objects cause mismatches of feature points, severely affecting localization accuracy. Semantic SLAM technology incorporates deep-learning-based semantic segmentation models to identify and filter out dynamic objects in the scene (e.g., workers, moving Automated Guided Vehicles (AGVs), lifts), retaining only static background features for pose estimation, significantly improving the robustness of localization and mapping [4]. Furthermore, semantic information can be used to construct semantic maps with object labels, providing high-level environmental understanding for subsequent task planning. For example, associating semantic labels such as "workbench," "conveyor belt," and "material bin" with corresponding regions in the map allows a robot to autonomously navigate and execute operations based on natural language commands like "go to the bin and pick up a Type-A workpiece."

2.4. Perception enhancement methods for complex manufacturing scenarios

To address difficult perception conditions such as high reflectivity, low texture, and featureless surfaces, researchers have proposed heterogeneous sensor fusion schemes. For instance, in narrow weld seams with high reflectivity, fusing vision and structured light information effectively overcomes reflective interference, enabling accurate identification and tracking of weld trajectories [2]. Structured light actively projects coded patterns to generate artificial features on surfaces lacking natural texture, thus solving the 3D reconstruction problem of low-texture objects. In large-component assembly scenarios, the combination of laser trackers and vision cameras enables coordinated coarse positioning over large ranges and fine local registration. Multimodal manufacturing data fusion perception is becoming a key technological direction for breaking through perception bottlenecks in complex environments; its core challenge lies in aligning differences in spatiotemporal resolution, coordinate systems, and noise characteristics among different sensors, as well as modeling semantic associations among heterogeneous information.

3. Flexible adaptive precision manipulation

3.1. Core contradiction and technical requirements of flexible manipulation

The core requirement for robotic manipulation in flexible manufacturing is to achieve a dynamic balance between flexible adaptability and high-precision manipulation [1]. On one hand, robots must adapt to workpieces of varying sizes, shapes, and materials—from electronic components of a few millimeters to structural parts several meters long, from rigid metals to flexible cables; on the other hand, precision assembly and micro-machining tasks require sub-millimeter or even micron-level operational accuracy. The contradiction between this broad workpiece adaptability and stringent precision requirements constitutes the fundamental technical challenge of flexible adaptive manipulation. Traditional position control methods have inherent deficiencies in contact force control and struggle to cope with dimensional tolerances and pose uncertainties of workpieces.

3.2. Deep-learning-based grasp detection and generalizable grasping

For grasping and manipulation, deep-learning-based grasp detection methods, by learning from large grasp datasets, can predict optimal grasp poses in real time, significantly improving generalizable grasping capabilities for diverse workpieces. Typical methods include: fully convolutional network-based grasp rectangle detection, which directly outputs grasp angles and widths for each pixel in an image; and point-cloud-based 6-DOF grasp pose estimation, suitable for grasping objects in cluttered piles. Researchers have also proposed grasp stability discriminative networks to evaluate the success probability of candidate grasp configurations. For family-type workpieces (e.g., identical part types of different sizes), transfer learning can substantially reduce the training data requirements for new workpiece grasp models.

3.3. Force/position hybrid control and precision assembly

Force/position hybrid control strategies are key technologies for precision assembly. Impedance control and admittance control dynamically regulate the force-position relationship, enabling robotic arms to possess compliant contact characteristics while maintaining positioning accuracy. The essential difference between them is: impedance control takes force input and outputs position corrections (suitable for position-controlled robots), while admittance control takes position input and outputs force commands (suitable for force-controlled robots). In typical tasks such as peg-in-hole assembly, force feedback guides the robotic arm to actively search for the hole position, effectively avoiding jamming and surface damage. Collaborative robots integrating tactile perception and deep policy learning have achieved micron-level accuracy in complex product assembly [5]. For soft robots, interaction technologies based on flexible sensing offer new solutions for compliant grasping [6]. Soft grippers utilize pneumatic or tendon-driven actuation to produce adaptive deformations, enabling damage-free grasping of irregular and fragile objects, though their load capacity and control precision still require improvement.

3.4. Progress in manipulation of deformable objects

Significant progress has also been made in robotic manipulation of flexible linear objects (cables, wire harnesses, hoses, etc.) [7]. The difficulties in manipulating such objects lie in: the objects themselves have no fixed shape, making modeling complex; they undergo continuous elastic or plastic deformation during operation; and the mapping between grasp points and operational goals is highly nonlinear. Existing methods include: vision-based state estimation for linear objects (describing object shape via skeleton extraction and B-spline fitting); learning-based grasping strategies (predicting suitable grasp points and operation sequences); and force-feedback-based insertion control (using force information to guide alignment when inserting connectors). However, stable manipulation of highly deformable objects still faces challenges such as complex theoretical models and large individual variations [7].

4. Intelligent decision-making for process planning and production line scheduling

4.1. Embodied intelligence requirements at the decision level

The value of embodied intelligence in flexible manufacturing is not limited to perception and manipulation but is even more manifested in the intellectualization of higher-level decision-making capabilities. The challenge of synergistic integration between general skills and specialized process knowledge requires industrial embodied intelligence to maintain task generalization capability while deeply adapting to the knowledge and constraints of specific manufacturing processes [1]. This means that a robot must not only know "how to do" (manipulation skills), but also "what to do," "when to do it," and "at which workstation to do it" (process planning and scheduling), and even adjust the original plan based on real-time status (dynamic replanning).

4.2. Deep reinforcement learning for dynamic scheduling

In process planning, production scheduling in flexible manufacturing systems is a typical NP-hard problem that requires efficient task allocation under equipment resource constraints. Traditional scheduling methods such as heuristic rules and mathematical programming often fail to respond in real time in dynamic environments. Deep Reinforcement Learning (DRL) offers new approaches for dynamic scheduling. Researchers have established multi-objective optimization models (balancing production cycle, equipment utilization, order priority, etc.) and combined them with DRL methods such as Deep Q-Network (DQN) and Proximal Policy Optimization (PPO) to propose real-time-data-driven dynamic intelligent scheduling mechanisms, enabling agents to adaptively select scheduling rules according to real-time workshop status (equipment load, work-in-progress inventory, emergency order insertion) [2]. Reward function design is crucial and requires trade-offs among different objectives.

4.3. Large-model-driven task planning and human-robot collaboration

Embodied intelligent manufacturing based on large models is becoming a frontier hotspot. Large-model-driven embodied intelligence focuses on how to combine the perception, reasoning, and logical capabilities of large models with embodied intelligence to improve the data efficiency and generalization capabilities of existing frameworks such as imitation learning and reinforcement learning [8]. Specifically, Large Language Models (LLMs) can be used to: decompose natural language instructions into executable sub-task sequences; retrieve relevant operation skills based on task descriptions; and generate alternative plans when encountering anomalies. Vision-Language Models (VLMs) can be used to understand object relationships and operation semantics in scenes. In specific applications, a research team from Tongji University proposed a control architecture with "decoupling of macroscopic semantic scheduling and microscopic visual execution," using an LLM as the global scheduling hub and visual foundation models as perception nodes, achieving dynamic task planning and precise execution for multi-form heterogeneous robots [2]. The advantage of this architecture is that the macroscopic scheduling layer does not concern itself with low-level control details, enabling rapid adaptation to different robot types; the microscopic execution layer focuses on visual servoing and force control, ensuring operational reliability.

4.4. Safe reinforcement learning and real-world deployment

Safe reinforcement learning is key to transferring intelligent decision-making from simulation to real robot systems [9]. In real manufacturing environments, decision errors may lead to equipment damage or personnel safety risks; therefore, researchers are devoted to applying reinforcement learning to robotic systems while guaranteeing safety. Major technical approaches include: constrained Markov decision process-based safe policy optimization (adding safety constraints to the optimization objective); barrier function-based methods (defining safe sets in the state space and ensuring policies do not leave the safe set); and environment projection-based methods (projecting unsafe actions to the nearest safe action). Despite considerable progress, the trade-off between safety and exploration efficiency remains the primary bottleneck limiting large-scale industrial application of reinforcement learning.

5. Development trends and challenges

5.1. Three-stage path of technological evolution

From the perspective of technological evolution, embodied intelligence for flexible manufacturing is advancing toward a three-stage direction of "cognitive enhancement—skill leap—system evolution" [10]. The first stage, "cognitive enhancement," focuses on improving the robot's environmental and task understanding capabilities, relying mainly on multimodal perception fusion and large-model reasoning. The second stage, "skill leap," aims to achieve the transition from single skills to skill composition and transfer, with the core being the construction of generalizable operation skill libraries. The third stage, "system evolution," concerns the autonomous evolutionary capability of the entire manufacturing system, enabling production lines to automatically reconfigure according to changes in production demand.

At the perception level, multimodal fusion and semantic SLAM technologies will further improve modeling capabilities for complex dynamic environments [4]; at the control level, manipulation methods based on deep learning and force/position hybrid control will promote the generalization of flexible grasping and precision assembly [5]; at the decision level, the integration of large models and safe reinforcement learning is expected to break through the bottleneck of transferring general skills to specialized processes [8, 9].

5.2. Current major challenges

Nevertheless, several important challenges remain. First, the scarcity of multimodal data hinders practical effectiveness; acquiring high-quality annotated data is costly. Data collection in industrial scenarios involves equipment downtime and privacy issues, and data distributions vary greatly across different processes, making sharing and reuse difficult. Second, perception understanding in complex manufacturing environments remains difficult, and AI hallucinations may pose application safety risks [8]. Large models may generate illogical or physically implausible steps in task planning, which is unacceptable in safety-critical applications. Third, software-hardware synergy issues constrain intelligence enhancement; the overall integration of embodied intelligence systems needs strengthening. Perception, decision, and control modules are often developed by different teams, with prominent problems in interface standards and communication latency. Fourth, the lack of ethical and legal frameworks brings standard compliance challenges; safety and privacy protection mechanisms for industrial embodied intelligence urgently need to be established [2]. For example, when a robot makes a wrong decision leading to an accident, how is liability determined? These issues have not yet been clearly regulated.

6. Conclusion

Centered on the closed-loop optimization framework of "perception–decision–execution," this paper systematically reviews the research progress of key technologies for robot embodied intelligence oriented toward flexible manufacturing. At the perception level, multimodal information fusion and semantic SLAM technologies are overcoming perception bottlenecks in complex manufacturing environments; at the control level, deep-learning-driven grasp detection and force/position hybrid control methods achieve a dynamic balance between flexible adaptability and high-precision manipulation; at the decision level, deep reinforcement learning and large-model technologies are driving the intelligent upgrading of process planning and production line scheduling. Embodied-intelligence-driven intelligent manufacturing is transitioning from theoretical exploration to practical application. With deeper technical breakthroughs and improved industrial ecosystems, it will provide key support for realizing a new generation of flexible, intelligent, and autonomous manufacturing systems.

It should be noted that this paper has certain limitations. Due to length and the review nature, this work primarily adopts qualitative analysis and has not conducted systematic quantitative comparisons or experimental evaluations of performance indicators for various key technologies, nor has it employed bibliometric methods for quantitative analysis of the evolution of research hotspots. In addition, this paper focuses on the technical aspects of perception, manipulation, and decision-making, and does not delve deeply into system-level engineering issues such as energy consumption, real-time performance, and software-hardware interface standards.

For future work, relevant research may further focus on the following directions: first, constructing high-quality multimodal datasets and simulation-to-reality transfer methods for industrial scenarios to alleviate the bottlenecks of data scarcity and high annotation costs; second, developing explainable and verifiable large-model task planning methods to suppress hallucinations and ensure reliability in safety-critical scenarios; third, promoting integrated design and standardized interfaces of perception–decision–control modules to improve system integration and real-time performance; and fourth, improving safety, privacy, and ethical-legal regulations for industrial embodied intelligence to provide institutional guarantees for large-scale deployment.

References

[1] Xu, K., Zhao, H., Hu, R., Yang, M., Liu, H., Zhang, H., & Yu, H. (2025). A review of embodied intelligence for flexible manufacturing. *Robot, 47*(4), 581–624. https://doi.org/10.13973/j.cnki.robot.250208

[2] Yuan, Y. G., He, Z. F., Li, W., & Hu, D. B. (2025). Development of intelligent manufacturing driven by embodied intelligence. *Strategic Study of CAE, 27*(3), 67–82.

[3] Lai, Y. N., Ye, X., & Ding, H. (2021). Research progress of major research plan on tri-co robots. *Journal of Mechanical Engineering, 57*(23), 1–11.

[4] Huang, Z. X., & Shao, C. L. (2023). Survey of visual SLAM based on deep learning. *Robot, 45*(6), 756–768.

[5] Wu, B. (2025). Research on flexible grasping and force control strategy of robot end-effector. *Mold Manufacturing*, (8), 222–224.

[6] Liu, W., Wang, Y., Duo, Y., Duan, J., Chen, X., Li, L., Liu, Y., & Wen, L. (2024). Recent progress of soft robot interaction based on flexible sensing. *Robot, 46*(2), 195–218. https://doi.org/10.13973/j.cnki.robot.230313

[7] Yu, M. R., & Li, X. (2024). Robotic manipulation of deformable linear objects: A survey. *Robot, 46*(5), 623–640.

[8] Bai, C. J., Xu, H. Z., & Li, X. L. (2024). Embodied-AI with large models: Research and challenges. *Scientia Sinica Informationis, 54*(9), 2035–2082.

[9] Zhang, C. X., Zhang, X. L., Xu, X., & Lu, Y. (2023). Safe reinforcement learning and its applications in robotics: A survey. *Control Theory & Applications, 40*(12), 2090–2103.

[10] Tao, Y., Wan, J. H., Wang, T. M., Xiong, Y. J., Wang, B. C., Zhang, W. B., Deng, C. Y., Tao, Y., Yang, G., & Wei, H. X. (2025). Establishing a new paradigm of embodied intelligence: A review of the current status and development trends in humanoid robot technology. *Journal of Mechanical Engineering, 61*(15), 121–147.