



REVIEW

Training Methods and Generation Technologies for Embodied Intelligent Robot Manipulation Skill Models: A Systematic Review

Lianpeng Li^{1,*}, Zhoujun Ruan¹, Zhichuang Wang², Haibo Zhang³, Hang Zhong⁴, Mingyang Li³
and Chunpeng Kang⁵

¹School of Automation, Beijing Information Science and Technology University, Beijing, China

²School of Artificial Intelligence, University of Science & Technology Beijing, Beijing, China

³Beijing Institute of Control Engineering, Beijing, China

⁴School of Artificial Intelligence and Robotics, Hunan University, Changsha, China

⁵Research and Planning Division, Information Center, Ministry of Agriculture and Rural Affairs, Beijing, China

*Corresponding Author: Lianpeng Li. Email: llp@bistu.edu.cn

Received: 16 April 2026; Accepted: 17 June 2026; Published: 13 August 2026

ABSTRACT: Endowing embodied intelligent robots with dexterous manipulation capabilities is paramount for executing complex, open-ended tasks. These capabilities are foundational to advancing true robotic autonomy, thereby facilitating precision assembly, seamless collaborative operations, and highly specialized maneuvers across diverse industrial and service sectors. Focusing on dynamic, unstructured environments where conventional programmed behaviors prove inadequate, this paper presents a systematic, quantitatively driven review of training methodologies and generation techniques for manipulation skill models within the domain of embodied artificial intelligence (AI). To provide rigorous trend validation, this study conducts a comprehensive bibliometric analysis and quantitative literature evaluation. By empirically mapping research distributions and methodological shifts across recent scholarly works, we substantiate our identification of mainstream research trajectories and developmental paradigms. This data-backed approach ensures the delineated landscape accurately reflects the consensus of the broader academic community, effectively mitigating subjective bias. Leveraging this validated framework, we rigorously analyze the core technical challenges currently hindering widespread deployment. These multifaceted challenges encompass high-dimensional nonlinear motion control, algorithmic sample inefficiencies, suboptimal reward formulations, the inherent semantic-to-physical alignment gap, and the computational constraints dictating large-scale applications. Furthermore, this review comprehensively examines key enabling technologies engineered to circumvent these barriers. Specifically, we systematically detail recent breakthroughs in motion control optimization, the evolution of advanced learning paradigms, complex task planning strategies, and synchronized multi-agent cooperative control frameworks. By critically evaluating these advancements, synthesizing the comparative efficacy of existing methodologies, and dissecting real-world deployment bottlenecks, this review establishes a robust theoretical foundation. Ultimately, it identifies practical, empirically supported pathways for future research to accelerate the transition of dexterous robotic manipulation from constrained laboratories to complex real-world applications.

KEYWORDS: Embodied intelligent robots; motion control; advanced learning methods; complex task planning; multi-agent collaboration

1 Introduction

As a critical interface for deep interaction between artificial intelligence and the physical world, embodied intelligent robots (EIRs) are physically situated systems capable of autonomous environmental perception, task cognition, and action execution [1,2]. Unlike traditional industrial robots confined to repetitive operations in structured settings, these systems actively adapt to dynamic environments to autonomously accomplish diverse tasks. This adaptability enables their deployment across complex, real-world scenarios, including domestic service, industrial maintenance, and emergency rescue. Consequently, they are driving a paradigm shift in the robotics industry, transitioning from traditional passive command execution to proactive perception and dynamic task adaptation.

Operating through a comprehensive perception-decision-control closed loop, embodied intelligent robots achieve an autonomous mapping from environmental inputs to physical actions. These systems utilize diverse sensors to extract multidimensional physical data, enabling robust environmental and task cognition. Based on this cognitive input and defined objectives, the robot formulates optimal motion trajectories and operational sequences, which actuators subsequently translate into physical movements. Concurrently, continuous real-time feedback during execution facilitates the dynamic refinement of decision-making and control strategies, ensuring cohesive and precise task completion.

Driven by advancements in multimodal perception [3], deep reinforcement learning, and foundation model-based planning [4], embodied intelligent robots now demonstrate preliminary proficiency in autonomous interaction and foundational task execution within unstructured environments.

The evolution of embodied robotics began in 1948 with Grey Walter's analog "electronic tortoises," which pioneered basic machine-environment interaction via simple photo-tactic behaviors (Fig. 1). By the 1970s, milestones such as Stanford's Shakey [5] and Waseda's WABOT-1 introduced symbolic logic planning and full-scale humanoid architectures, respectively. The late 1980s saw a shift toward Rodney Brooks's behaviorist paradigm of "intelligence without representation," paving the way for platforms like Honda's ASIMO [6] to establish kinematic standards for bipedal locomotion in 2000. Subsequent systems—including iCub [7], PR2, and Boston Dynamics' hydraulic Atlas—significantly advanced cognitive developmental robotics, ROS standardization, and highly dynamic mobility. Entering the 2020s, the integration of foundation models (e.g., Google RT-2) and advanced humanoid platforms (e.g., Tesla Optimus) has expanded the capacity for semantic understanding, addressing previous limitations in perception and task generalization. Since 2024, the emergence of platforms such as Fig. 1, 1X NEO, and the fully electric Atlas indicates a progressive transition from controlled laboratory settings to real-world application scenarios. Driven by end-to-end learning architectures, current embodied AI research focuses on enhancing generalizability, scalable deployment, and integration into unstructured human environments.

The deployment of embodied intelligent robots, characterized by advanced autonomy and sophisticated manipulation capabilities, has rapidly proliferated across industrial manufacturing [8], healthcare [9], and specialized operations [10] (Fig. 2). Catalyzed by the advent of the Industry 5.0 paradigm, contemporary manufacturing and service sectors are undergoing a profound transition: from the efficiency-centric, rigid automation characteristic of Industry 4.0, toward a resilient, sustainable, and fundamentally human-centric model of human-robot symbiosis [8]. This transformative paradigm mandates highly agile systems capable of facilitating mass personalization and dynamically adapting to pervasive environmental uncertainties. Within this vision, embodied agents must transcend passive, deterministic execution. To effectively share unstructured workspaces and autonomously assume physically demanding or hazardous duties—thereby emancipating human operators for cognitive and value-added roles—these systems require near-human levels of dexterous manipulation and cognitive plasticity. Consequently, they are increasingly tasked with executing high-precision, contact-rich operations, encompassing complex assembly [11], fine-grained

manipulation [12], and close-proximity collaboration alongside human counterparts [13]. However, conventional skill synthesis frameworks exhibit acute vulnerabilities regarding adaptability, precision, and operational efficiency when confronted with the stochastic dynamics of real-world environments and the stringent safety constraints of human-centric operations. Addressing these critical bottlenecks through advanced training methodologies and novel policy generation techniques has thus emerged as a paramount research priority. Surmounting these limitations is essential for transitioning robotic capabilities from deterministic automation to resilient, dynamically adaptive intelligence.

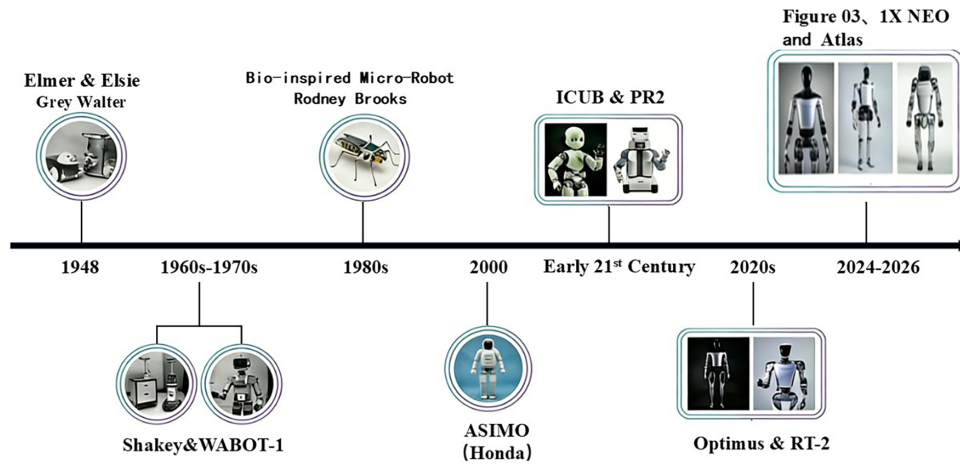


Figure 1: The evolutionary trajectory of embodied intelligent robotics.

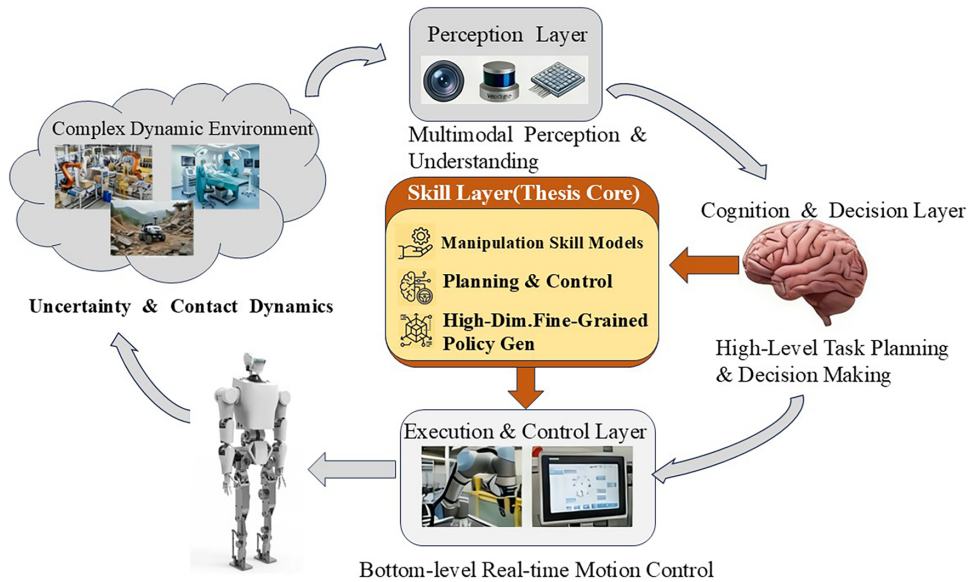


Figure 2: General system architecture and positioning of manipulation skill models.

Global robotics competition has decisively shifted from hardware-centric design to underlying AI technologies. While traditional giants (e.g., ABB, FANUC) adapt their deterministic automation expertise to embodied AI, emerging pioneers are driving breakthroughs through heterogeneous paradigms. For instance, Tesla Optimus pursues generalized physical AI via vision-only, end-to-end neural networks,

whereas Figure AI leverages multimodal foundation models for advanced reasoning and task planning. Concurrently, Chinese enterprises are rapidly maturing proprietary ecosystems, companies like Unitree utilize large-scale reinforcement learning and Sim-to-Real transfer to optimize bipedal control, while Agibot pioneers standardized robotic operating systems and data architectures (e.g., Dexterous OS). Collectively, these industry-wide advancements underscore a rapid acceleration toward highly capable manipulation skill models.

Advanced manipulation models require the fusion of efficient motion control, task planning, intelligent learning, and multi-agent coordination mechanisms. This convergence is essential for delivering stable, precise, and generalizable performance in unstructured environments, highlighting that empowering embodied robots with human-like fine manipulation skills and dynamic adaptability remains a critical research challenge.

In response to these challenges, embodied AI research has established a multidimensional, cross-hierarchical technological framework. This review systematically examines this framework through four core directions: motion control optimization to ensure execution precision, intelligent learning methods to enhance generalization, complex task planning to resolve long-horizon decision-making bottlenecks, and multi-agent collaboration to expand operational boundaries. Drawing on recent global advancements, we thoroughly investigate methodological innovations, mechanism optimizations, and engineering applications across these domains. By analyzing current technical bottlenecks and potential breakthrough trajectories, this paper outlines future trends for dexterous manipulation skill models, ultimately providing a comprehensive reference for ongoing research and engineering practice.

To ensure methodological rigor and transparency, the literature selection process was conducted in strict adherence to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. During the initial identification phase, a systematic search was executed across three prominent academic databases. Following a stringent pipeline involving deduplication, content screening, and comprehensive full-text eligibility evaluation, a total of 107 relevant publications were formally curated and incorporated into the core analytical scope of this review. As illustrated in [Fig. 3](#), PRISMA checklists are available in the Supplementary Materials.

In this review, two authors independently screened titles and abstracts, then reviewed full texts of the remaining papers, resolving disagreements through discussion. Data extraction was also performed independently by two authors using a standardized form, focusing on robotic platforms, algorithmic frameworks, and application scenarios; missing data were not sought from the original investigators. No automation tools were used in either screening or extraction. Studies were qualitatively grouped based on the core robotic tasks they addressed (e.g., perception, path planning, control). The technical characteristics, advantages, and limitations of the methods were organized into comparative tables, and conceptual diagrams were created to visually illustrate their relationships.

A limitation of our review process is that the literature search was mostly English-language publications, potentially excluding relevant non-English literature or industry technical reports.

During the preparation of this work, the authors used AI tools solely for the purpose of language polishing, grammar correction, and improving the readability of the manuscript. No AI tools were used for literature summarization, data analysis, figure generation, or drafting the core scientific content of the paper. After using these tools, the authors reviewed and edited the content as necessary and take full responsibility for the accuracy and integrity of the final publication.

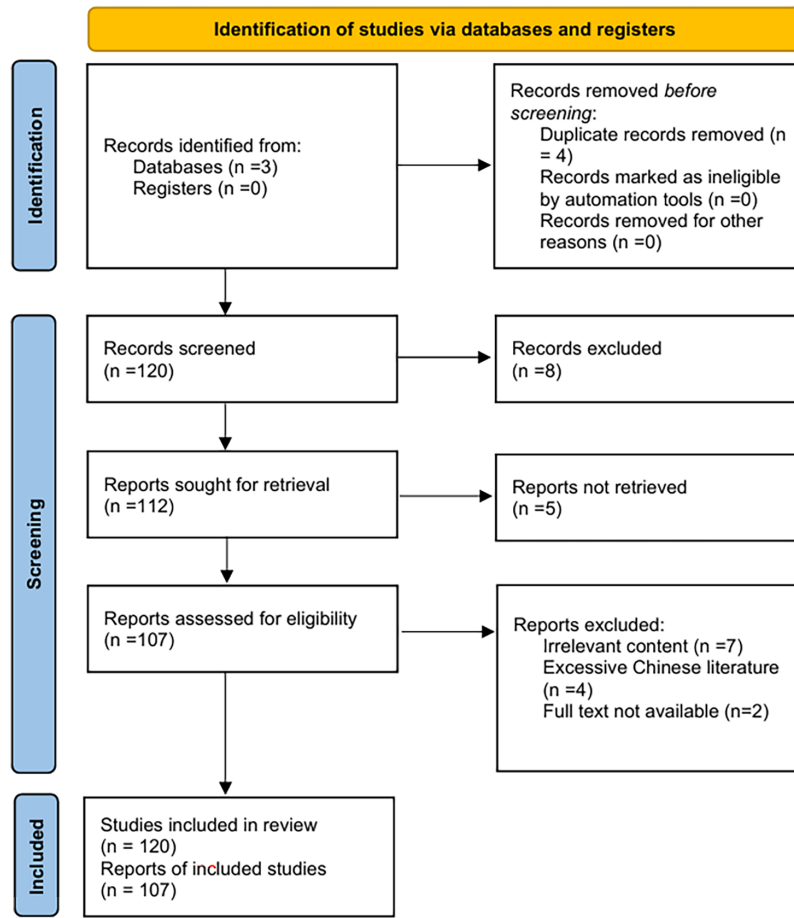


Figure 3: PRISMA flow diagram of the study selection process.

2 Motion Control Optimization for Embodied Robots

Motion control optimization is fundamental to achieving dexterous and fine manipulation. Through algorithmic innovation, sensorimotor integration, and refined reward mechanisms, this field advances robotic precision, stability, and environmental adaptability. Its ultimate goal is to enable autonomous execution under complex constraints, whereby robots generate optimal control commands to fulfill predefined tasks with high efficiency and adaptability.

2.1 Classical Control Methods

Classical control remains fundamental to robotic motion control, providing a robust and empirically validated theoretical framework for addressing deterministic control problems. Pivotal methodologies such as Proportional-Integral-Derivative (PID) control [14], Nyquist's frequency response analysis [15], and Evans' root locus method [16] collectively define the analytical arsenal of classical control theory. Notably, the root locus method remains actively utilised in contemporary biomechanical and mobile robotic applications, such as human gait analysis [17].

For conventional tasks such as manipulator trajectory tracking and mobile robot path planning, classical control methods yield consistently stable performance. In the frequency domain, for example, researchers have applied the Frequency Response Function (FRF) to predict the multimodal dynamics of milling robots, effectively mitigating stability issues during variable posture machining [18]. Furthermore, the classical PID

architecture remains foundational in mobile robotics; notably, Karthik et al. [19] developed a dual-mode PID controller that enables high-precision, encoder less navigation in warehousing environments. Additionally, the root locus method continues to serve as a crucial tool for analyzing closed-loop pole distributions to ensure robustness in complex setups like human-robot collaboration. Ultimately, these applications underscore the enduring efficacy of classical techniques, particularly for Single-Input Single-Output (SISO) Linear Time-Invariant (LTI) systems and specific nonlinear extensions.

2.2 Modern Control Methods

Addressing the limitations of classical control in complex environments, modern methods—anchored in state-space modeling [20–22]—provide a comprehensive framework tailored for Multiple-Input Multiple-Output (MIMO), nonlinear, and time-varying systems. This paradigm tackles specific operational challenges through targeted strategies: Kalman filtering [23,24] for precise state estimation, Model Predictive Control (MPC) for constrained time-domain optimization and sliding mode and adaptive control to mitigate nonlinearities and parameter uncertainties. This precise characterization of system dynamics offers significant advantages in demanding applications, such as collaborative robot interaction, flexible manipulator control [25], and mobile robot decision-making [26]. Ultimately, by transcending the assumptions of Linear Time-Invariant (LTI) systems, modern control lays a robust foundation for advancing robotics toward higher-dimensional complexity and intelligence.

Table 1 summarizes the evolution of motion control optimization for tracked robots. Early research primarily targeted travers ability over specific terrains by formulating fundamental geometric, kinematic, and dynamic models. For structured scenarios like stairs, several studies [27–29] proposed sequential dynamic models relying on manual operational frameworks. These methods plan flipper articulation by conforming to the terrain envelope and derive optimal postures from energy stability margins. However, these models remain strictly limited to simple, uniform terrains. When deployed in complex, real-world environments with superimposed obstacles or dynamic terrain variations, such approaches frequently prove inadequate and are highly susceptible to control failure.

Table 1: History of motion control optimization for tracked robots.

Evolutionary Stage	Core Methodology	Key Mechanisms	Advantages	Limitations
Early Research [27–29]	Kinematic & Dynamic Modeling	Sequential motion dynamic models	Established foundational travers ability theory. Resolved single-obstacle crossing.	Poor environmental adaptability. Prone to failure. limited to simple terrains.
Recent Advances [30]	Deep Reinforcement Learning (DRL) & Visual Perception	CNN-based feature extraction & DDPG. Re-al-world da-ta-driven algorithms	Accurately models non-linear contact dynamics. Enhances adaptability to complex terrains.	vulnerable to visual noise, low sample efficiency during early exploration.
State-of-the-Art [31–33]	Low-Cost Simulation & Discrete-Space Algorithms	Pymunk simulation environment. Discrete-space AT-D3QN algorithm	Significantly reduces training thresholds. achieves superior obstacle-crossing speed and stability.	Addresses critical bottlenecks in cost and training efficiency.

Addressing the adaptability bottlenecks of analytical models, recent machine learning advancements have catalyzed a shift toward learning-based control. Early end-to-end Deep Reinforcement Learning (DRL) frameworks [30] combined Convolutional Neural Networks (CNNs) for visual extraction with the Deep Deterministic Policy Gradient (DDPG) algorithm for policy optimization. Despite their pioneering nature, these methods suffered from high sample complexity, prohibitive computational costs, and control instabilities caused by visual perturbations. Subsequent research improved stability via reward shaping [31,32] and enhanced terrain adaptability by integrating real-world data into Q-learning [33]. Nevertheless, these approaches remain constrained by sub-optimal sample efficiency and limited generalization across diverse scenarios, leaving the full potential of learning-based algorithms unrealized.

Recent breakthroughs feature the development of a cost-effective Pymunk simulation environment [34] alongside the AT-D3QN autonomous control algorithm, tailored specifically for discrete action spaces. This approach not only facilitates highly efficient obstacle traversal but also significantly outperforms manual teleoperation in both speed and stability. Consequently, it effectively mitigates the economic barriers that have historically hindered the real-world deployment of such systems.

In the realm of legged locomotion (Fig. 4), researchers have formulated targeted solutions for scenario-specific challenges. To address the complexities of highly dynamic motion retargeting, for instance, Grandia et al. [35] introduced a generalized framework based on differentiable optimal control. This bi-level optimization scheme parameterizes the system to reconcile discrepancies between source inputs and robot dynamics. Specifically, the inner loop leverages Model Predictive Control (MPC) for optimal trajectory generation, while the outer loop minimizes tracking errors to refine these parameters. Furthermore, by integrating projection techniques with Matrix Riccati equations to resolve complex constraints, this approach successfully balances physical feasibility with perturbation robustness, thereby establishing a practical paradigm for multi-constrained dynamic control.

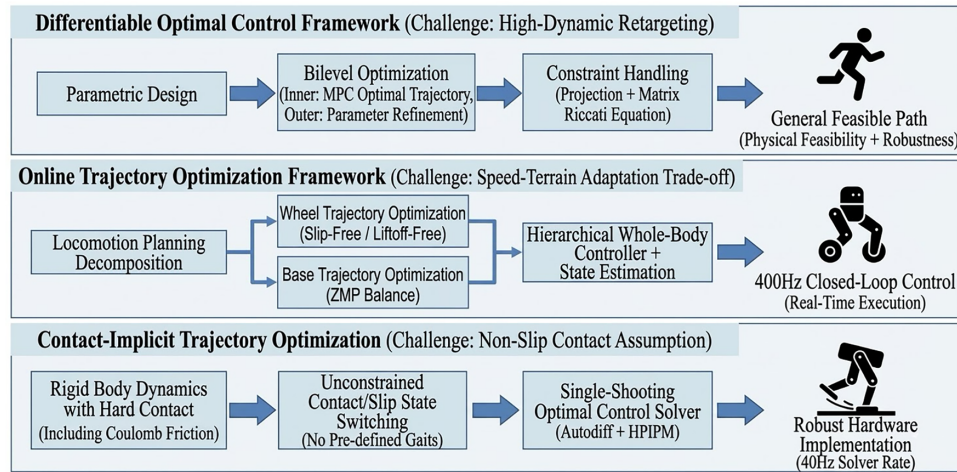


Figure 4: Legged robots evolution of control strategies from trajectory planning to contact mechanisms.

To address the inherent speed-stability trade-off in wheeled-legged robots, Bjelonic et al. [36] introduced a hierarchical online trajectory optimization framework. This approach decouples motion planning into base posture optimization and wheel trajectory generation, explicitly accounting for both stance and flight phases. By integrating a Whole-Body Controller (WBC), the system achieves 400 Hz closed-loop control, successfully facilitating agile maneuvers across complex terrains.

At the fundamental level of contact mechanics, Carius et al. [37] introduced a contact-implicit trajectory optimization method to relax the conventional no-slip assumption. By directly embedding frictional contact models into the system dynamics, this approach enables autonomous transitions between contact and slip states without relying on predefined gait sequences. Powered by efficient solvers, this real-time optimization yields trajectories suitable for direct hardware deployment, thereby significantly enhancing locomotion robustness in complex environments.

2.3 Motion Control Optimization for Embodied Intelligence

Whereas modern control focuses on optimizing performance indices subject to multivariable constraints, embodied AI tightly couples perception and decision-making. Its primary goal is to achieve broad task generalization and profound semantic understanding in complex environments. Driven by generative AI and large-scale simulations, the field is shifting from rule-based to learning-based paradigms. Ultimately, by leveraging massive datasets, this approach endows robots with human-like operational capabilities, facilitating autonomous interaction within the physical world.

Traditional imitation learning heavily relies on Mean Squared Error (MSE) for action regression, fundamentally failing to capture multimodal distributions. Consequently, when an agent confronts equally viable alternatives—such as turning left or right—MSE averages these modes, producing an interpolated trajectory that inevitably triggers collisions or deadlocks. To circumvent this, Diffusion Policy introduces the generative denoising paradigm into robot control, formulating action generation as an iterative process that synthesizes coherent action sequences from pure Gaussian noise.

This denoising paradigm empowers embodied agents to resolve complex tasks featuring multiple viable solutions. Diffusion Policy not only precisely models non-Gaussian, multimodal action distributions enabling the robot to commit to a single execution mode, but also ensures temporal consistency by predicting future action sequences. Consequently, it effectively mitigates the high-frequency jitter typically associated with standard neural network policies.

Advancing this paradigm, Ren et al. [38] introduced Diffusion Policy Policy Optimization (DPPO), effectively integrating diffusion models with reinforcement learning. By eschewing uninformed isotropic noise, DPPO leverages the representational capacity of diffusion models to drive structured exploration. Consequently, in long-horizon, sparse-reward assembly tasks, this framework exhibits more robust training convergence and superior Sim-to-Real transferability than conventional Gaussian policies.

Although Diffusion Policy effectively resolves temporal consistency and multimodality in action generation, achieving general-purpose intelligent control necessitates that embodied agents comprehend complex semantic instructions and perform long-horizon planning. This requirement has catalyzed a profound convergence between foundation models and low-level control systems.

Vision-Language-Action (VLA) models [39], exemplified by systems like Google's RT-2 and NVIDIA's GR00T [40], represent a significant shift in cross-modal representation learning. Architecturally, these frameworks map high-dimensional visual inputs and natural language instructions into a shared latent space, subsequently discretizing continuous robotic actions into text-like tokens. By unifying these modalities within a single Transformer objective, Vision-Language-Action (VLA) models establish a centralized "robotic brain" capable of exploiting internet-scale pre-training. This paradigm significantly enhances semantic reasoning and zero-shot manipulation of novel objects, pushing the generalization boundaries of robotic control.

Despite these architectural advancements, the practical deployment of VLA models is severely constrained by critical bottlenecks. From a dataset perspective, while VLAs benefit from abundant internet-scale

text and images, high-quality and diverse robotic action data remains intrinsically scarce. This “embodiment data gap” restricts the model’s ability to ground semantic knowledge into precise physical dynamics. Furthermore, the massive parameterization inherent to these foundation models introduces substantial inference latency. The autoregressive nature of token generation in large Transformers precludes the high-frequency control loops (typically > 500 Hz) required for dynamic stability, rendering direct, real-time reactive control computationally prohibitive on current edge hardware.

To mitigate logical inconsistencies in long-horizon tasks, contemporary research has integrated Chain-of-Thought (CoT) reasoning [41] into the embodied control loop, transitioning from black-box action regression to explicit cognitive inference. To circumvent the aforementioned inference bottlenecks, current architectural trends widely adopt a hierarchical decoupling paradigm [42]. In this structure, the VLA model serves as a low-frequency, high-level planner to orchestrate task objectives or keyframes, while a lower-level, high-frequency execution policy—such as a Whole-Body Controller (WBC) or a lightweight diffusion model—handles reactive continuous control.

While this hierarchical framework represents a promising approach to bridging foundation models with traditional control, it remains fraught with systemic vulnerabilities in robustness, latency, and safety. The asynchronous operation between the low-frequency planner and high-frequency executor often introduces latency-induced instabilities in dynamic environments. Moreover, the abstraction of task objectives creates a “semantic-to-control gap,” where high-level plans may fail to account for fine-grained geometric or dynamic constraints, leading to cascading execution errors. Most critically, the inherent unpredictability—such as hallucinations of large-parameter planners poses severe safety risks when translated into physical torque commands. Therefore, achieving deterministic stability and seamless, safe integration across these hierarchical layers remains a profound challenge in current research, rather than a resolved reconciliation.

Due to their anthropomorphic design, humanoid robots offer distinct morphological advantages over quadrupedal and wheel-legged platforms in navigating human-centric environments and executing versatile tasks. However, they also confront formidable challenges arising from high-dimensional nonlinearity and under-actuated dynamics, necessitating a transition in control strategies from model-driven to data-driven paradigms. Within the model-based framework, the integration of Model Predictive Control (MPC) [43] and Whole-Body Control (WBC) [44] has become the mainstream, effectively ensuring dynamic stability by incorporating angular momentum constraints and contact force optimization, yet, this approach is often hampered by modeling inaccuracies in unstructured terrains. In contrast, leveraging large-scale parallel simulation [45] and domain randomization [46] has not only bypassed the computational complexities of contact dynamics to achieve robust locomotion, but also catalyzed the emergence of integrated mobile manipulation—coupling lower-body locomotion with upper-body dexterous execution. This synergy establishes a critical technical foundation for autonomous embodied agents to achieve truly versatile operations in complex, real-world scenarios.

In conclusion, the evolution of robotic motion control optimization indicates that advancing embodied intelligence to higher levels of autonomy requires transcending isolated technical domains to foster a tripartite synergy between theory, engineering, and application scenarios. Future efforts must prioritize narrowing the sim-to-real gap by ensuring the hardware-aware adaptation of algorithms to physical constraints. Furthermore, the development of modular, hardware-agnostic optimization frameworks will be essential to mitigate R&D overhead and facilitate seamless adaptation across heterogeneous robotic platforms.

3 Skill Training and Learning Methodologies for Embodied AI

3.1 Core Reinforcement Learning Methodologies for Skill Synthesis

The integration of Reinforcement Learning (RL) into embodied robotics represents a methodological shift from deterministic, pre-programmed control to data-driven decision-making. As illustrated in Fig. 5, RL optimizes control policies through continuous environmental interaction [47,48] and iterative trial-and-error exploration [49]. By reducing the reliance on explicit heuristic rules and precise analytical models, RL provides a viable framework for managing high-dimensional and nonlinear motion dynamics. However, while RL enables the synthesis of complex behaviors, its practical deployment in physical systems is frequently hindered by critical limitations, including severe sample inefficiency, the complexity of reward engineering, and the sim-to-real gap. Consequently, achieving robust and generalized adaptation in dynamic, non-stationary environments remains a persistent challenge.

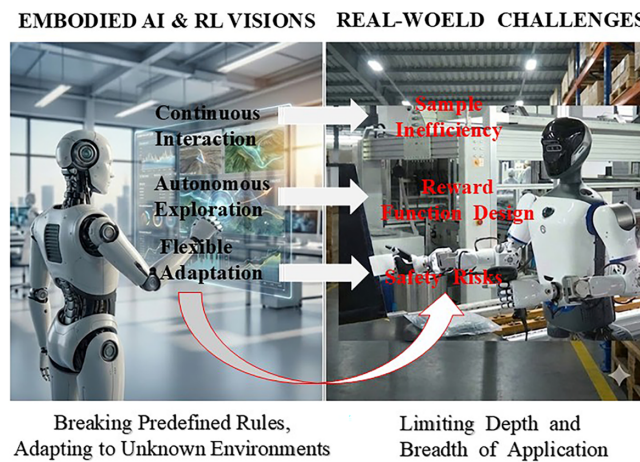


Figure 5: Reinforcement learning for embodied intelligent robots: vision and real-world challenges.

Despite its theoretical promise, the deployment of Reinforcement Learning (RL) in embodied systems is impeded by several fundamental bottlenecks. Foremost, extreme sample inefficiency renders the training process computationally prohibitive, often requiring millions of environmental interactions to synthesize even elementary skills. Furthermore, formulating reward functions that perfectly balance task maximization with implicit physical constraints remains a notoriously intractable engineering challenge. Additionally, the inherent stochasticity of physical environments introduces severe safety risks during the trial-and-error exploration phase. Ultimately, the sim-to-real gap acts as the primary barrier to real-world deployment, frequently causing policies optimized in simulation to fail in practice. Together, these deep-seated impediments severely constrain the scalability of RL in practical, high-stakes domains, such as large-scale industrial manufacturing [50] and hazardous operations.

To overcome these practical bottlenecks, targeted algorithmic innovations in Reinforcement Learning (RL) have emerged as a critical focal point across academia and industry. This paradigm shift necessitates algorithmic advancements that simultaneously enhance sample efficiency, refine reward shaping mechanisms, and embed robust safety constraints directly into the control architecture. Concurrently, the formidable technical hurdles of bridging the sim-to-real gap must be surmounted. Ultimately, only through synergistic breakthroughs in theory and engineering—coupled with pioneering applications in high-value scenarios like dexterous manipulation and dynamic environmental adaptation—can the evolution of RL in embodied agents be genuinely propelled forward.

3.1.1 Additive Linear Reward Summation

Embodied agents acquire skills through continuous environmental interaction, extracting generalized competencies to execute complex tasks [51,52]. However, as environmental dynamics and the kinematic complexity of these systems escalate, the difficulty of policy optimization grows exponentially, necessitating explicit constraints to guide the learning process. Relying on the principle of cumulative reward maximization [53,54], Reinforcement Learning (RL) typically formulates these operational constraints as constituent terms within the reward function. While this approach has demonstrated efficacy in demanding domains like legged locomotion and dexterous fine manipulation, it harbors a fundamental flaw. Conventional RL frameworks generally rely on a rudimentary linear scalarization of reward components. This formulation compels the agent to optimize competing objectives simultaneously, frequently inducing gradient interference and severe policy oscillation, which substantially degrade overall learning efficiency [55].

3.1.2 Hybrid Dynamic Policy Gradient Method

To resolve the core challenge of multi-objective conflicts, researchers have progressed from static decoupling to dynamic architectural optimization. The Hybrid Reward Architecture (HRA) [56] employs a “divide-and-conquer” strategy: by decomposing the reward function and mapping each reward component to a dedicated value function, it suppresses cross-objective interference and markedly improves learning efficiency. Building upon this, subsequent work introduced the Hybrid Dynamic Policy Gradient (HDPG) method. Built as an extension of the Deep Deterministic Policy Gradient (DDPG) algorithm [57], HDPG produces a vectorized value output $Q(s, a)$ as formulated in Eq. (1):

$$Q(s, a) = [Q_1(s, a), Q_2(s, a), \dots, Q_K(s, a)]^T \quad (1)$$

where K is the number of reward components, and each corresponds to a specific physical constraint reward. This method maintains decoupled Q values, thereby enabling the calculation of gradient suggestions from each sub-objective regarding the action. Let the parameters of the Actor network be θ^π and those of the Critic network be θ^Q . The gradient contribution k of each sub-objective ∇J_k can be derived via the chain rule, as shown in Eq. (2):

$$\nabla J_k(\theta^\pi) = \mathbb{E}_{s \sim \rho^\pi} \cdot \vartheta \quad (2)$$

$$\vartheta = [\nabla_a Q_k(s, a | \theta_k^Q) |_{a=\pi(s)} \cdot \nabla_{\theta^\pi} \pi(s | \theta^\pi)] \quad (3)$$

To enhance policy optimization, a dynamic priority mechanism is implemented by introducing the dynamic weight vector m_k , as formulated in Eq. (3), which modulates the contribution of each reward branch. Unlike predefined hyperparameters, these weights m_k are adaptively determined on-the-fly, based on empirical statistics accumulated during the training process. The resulting hybrid gradients are then expressed as shown in Eqs. (4) and (5).

$$m_k = \frac{K(\mu_k + e^{\sigma_k^2})}{\sum_{j=1}^K (\mu_j + e^{\sigma_j^2})} \quad (4)$$

$$\nabla_{\theta^\pi} J = \sum_{k=1}^K m_k \nabla J_k \quad (5)$$

This mechanism aims to steer embodied agents toward mastering fundamental skills before progressively tackling high-complexity challenges, thereby manifesting a curriculum-based, incremental learning

effect akin to human cognition. Nevertheless, such approaches encounter inherent bottlenecks; specifically, the dynamic weighting rules rely heavily on hand-crafted heuristics and lack sufficient autonomy. This reliance ultimately compromises the algorithm's robustness and generalization when deployed in novel or unstructured environments.

3.1.3 Automated Hybrid Reward Scheduling Framework

To transcend the constraints inherent in manual heuristic design, recent frontier studies have begun investigating the potential of Large Language Models (LLMs) to facilitate automated rule synthesis. A pivotal advancement in this domain is the Automated Hybrid Reward Scheduling (AHRS) framework [58], which instantiates this automated approach. The conceptual architecture is illustrated in Fig. 6.

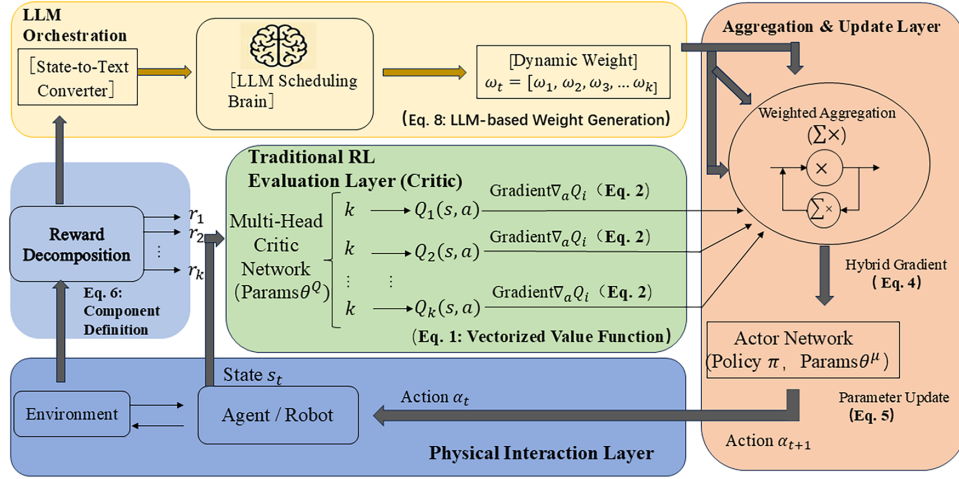


Figure 6: Schematic diagram of the automated hybrid reward scheduling (AHRS) framework challenges.

The framework begins by decomposing the total reward function R into K distinct physical components r_k . Consequently, the aggregate reward t at time step r_t is formulated as a weighted linear combination of these constituent elements:

$$r_t = \sum_{k=1}^K w_t^{(k)} \cdot r_t^{(k)} \quad (6)$$

Here, $w_t^{(k)}$ represents the relative weight ascribed to the k -th reward component within the current operational phase. Leveraging this decomposition, the framework instantiates a multi-branch value network architecture, where the state-action value function $Q(s_t, a_t; \theta)$ is approximated as a linear combination of the individual value functions derived from each respective branch:

$$Q(s_t, a_t; \theta) = \sum_{k=1}^K w_t^{(k)} \cdot Q_k(s_t, a_t; \theta_k) \quad (7)$$

Subsequently, the framework introduces a Large Language Model (LLM) to serve as a sophisticated orchestration engine. By harnessing the LLM's formidable semantic comprehension and logical reasoning capabilities, the system can autonomously retrieve optimal strategies from the rule buffer S_t based on

the formalized training state description $\mathcal{B}$. This process culminates in the adaptive computation of the importance weight vector w_t across all reward components:

$$w_t = \text{LLM}(\mathcal{S}_t, \mathcal{B}, \mathcal{P}) \quad (8)$$

In this context, $\mathcal{P}$ functions as the guiding prompt. via this orchestration, the LLM directs the optimization trajectory of the policy π across the parameter space φ to maximize the objective function. The proposed framework effectively bridges the gap between manual heuristic design and the intelligent, autonomous synthesis of weighting rules. Consequently, this significantly bolsters both the autonomy and the generalization proficiency of embodied agents during the acquisition of complex skills.

In the field of reinforcement learning for bipedal robots, researchers have proposed various tailored solutions through architectural innovation and policy optimization to address core bottlenecks across diverse scenarios, such as complex skill acquisition, path tracking, and gait generation.

3.1.4 Hierarchical Control and Curriculum Learning Strategies

Addressing the coupled challenges of complex skill acquisition and the sim-to-real gap, Haarnoja et al. [59] proposed a two-stage Deep Reinforcement Learning (DRL) pipeline. Incorporating policy distillation, domain randomization, and external force perturbations, this approach enabled the robust transfer of agile locomotion skills to the OP3 robot. For wheeled-bipedal platforms, Zhu et al. [60] devised a three-tier hierarchical control architecture coupling the Deep Deterministic Policy Gradient (DDPG) algorithm with conventional PID controllers. This hybrid paradigm obviates exhaustive manual tuning, substantially augmenting the path-tracking robustness of the Intelligent Ground Operations Robot (IGOR) robot.

Beyond architectural optimizations, decoupling policy learning from predefined reference trajectories remains a pivotal frontier in locomotion research. Addressing this, Siekmann et al. [61] devised a periodic reward composition framework. By formulating a phase-conditioned reward function and synergizing Proximal Policy Optimization (PPO) with dynamics randomization, this approach synthesizes diverse gaits devoid of explicit kinematic priors. Ultimately, this paradigm endows the Cassie biped with the capacity for seamless multi-gait transitions and robust zero-shot transferability to unstructured outdoor terrains.

Advancing the stability and sample efficiency of omnidirectional locomotion, Rodriguez and Behnke [62] introduced a curriculum-driven Deep Reinforcement Learning (DRL) framework (Fig. 7). By orchestrating a velocity scheduler to progressively escalate task complexity, this method utilizes nominal postural priors to guide the optimization trajectory. To preclude unsafe explorations, a Beta distribution policy strictly bounds the action space. Coupled with a composite reward function—targeting velocity tracking and postural regularization—and stochastic noise injection, this systematic paradigm empowered the Nimbro-OP2X biped to synthesize robust multidirectional gaits. Consequently, the learned policies exhibited exceptional perturbation resilience, ensuring seamless sim-to-real deployment on physical hardware.

For a systematic evaluation of the three representative reward paradigms, Table 2 comprehensively characterizes traditional linear scalarization, hybrid architectural decoupling (HDPG), and LLM-driven automated scheduling (AHRS) in terms of their core design, computational complexity, and performance boundaries.

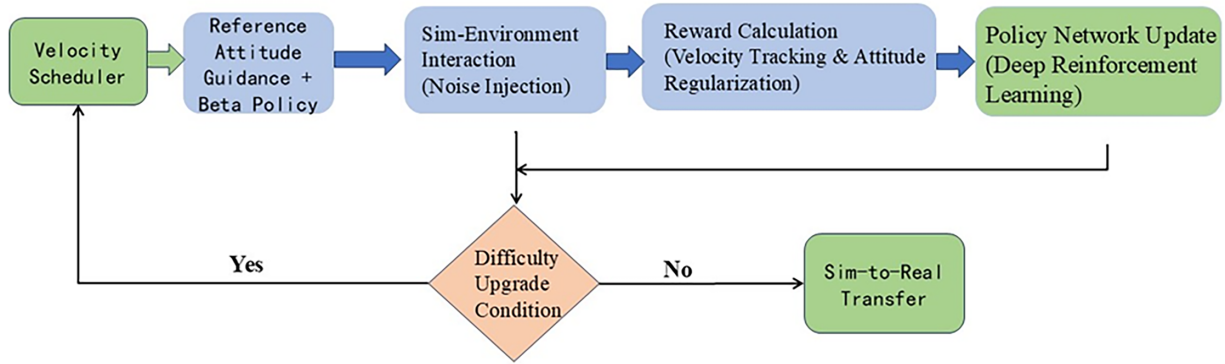


Figure 7: Flowchart of curriculum-driven reinforcement learning training for humanoid robots.

Table 2: Evolution of reward mechanisms in reinforcement learning for embodied AI.

Evolutionary Stage	Representative Methods	Core Mechanisms & Strategies	Computational Complexity	Major Advantages	Limitations
Traditional Baseline	Standard Reinforcement Learning [55]	Formulates constraints as reward components via linear scalarization; compels the agent to optimize all objectives concurrently within a single time horizon.	Baseline reference (100 percent). Minimal overhead at 1.0 times FLOPs. Utilizes standard architectures, maintaining low memory footprint and normal inference latency.	Mature theoretical foundation; proven effective in specific tasks such as legged locomotion and dexterous manipulation.	Parallel learning easily triggers oscillations in optimization direction; struggles with high-dimensional, multi-constraint complex tasks.
Architectural Innovation	Hybrid Reward Architecture and HDPG [56]	Decouples reward functions by matching each component with a dedicated value function; (HDPG) adjusts weights to implement progressive learning from simple to complex.	250 to 400 percent VRAM increase; 150 to 200 percent longer training time. Parameters scale linearly (three to five times) due to the parallel maintenance of multiple decoupled value networks.	Avoids direct conflict between multiple objectives; the learning process is more targeted and hierarchical.	Dynamic weight rules are highly dependent on manual design; lacks adaptive capabilities and is difficult to transfer to new scenarios.
Intelligent Break-through	Automated Hybrid Reward Scheduling [59]	Leverages LLM reasoning capabilities to filter adaptation rules; autonomously calculates importance weights for each reward component to guide optimization.	Over 10,000 percent computational surge. Iterative Large Language Model (LLM) inference increases FLOPs by multiple orders of magnitude, introducing critical real-time latency bottlenecks.	High Generalization: Significantly improves adaptability in learning complex skills.	Lack of interpretability in weight computation. Over-reliance on LLM priors for rule generation.

3.2 Core Methods for Intelligent Training

In embodied AI research, while advanced control architectures and Reinforcement Learning (RL) paradigms establish robust algorithmic frameworks for skill acquisition, procuring high-fidelity training data

and constructing generalizable perceptual representations remain the primary bottlenecks impeding cross-task generalization. Conventional training pipelines consistently suffer from extreme sample inefficiency, cost-prohibitive physical interactions, and a fundamental deficit in semantic reasoning. To circumvent these limitations, the community is converging on two complementary trajectories. First, Augmented Reality (AR) and Virtual Reality (VR) technologies are leveraged to engineer immersive, human-in-the-loop (HITL) training environments, injecting human cognitive priors to resolve the intractable challenges of demonstrating and guiding complex tasks. Second, large-scale visual pre-training models are utilized to distill universal physical priors and semantic features from massive multimodal datasets, endowing embodied agents with a robust perceptual foundation for compositional generalization. Consequently, this section systematically reviews recent advancements across these two pivotal dimensions: AR/VR-assisted skill acquisition and visual pre-training methodologies.

3.2.1 Skill Acquisition via Augmented and Virtual Reality Training

Integrating Augmented Reality (AR) and Virtual Reality (VR) establishes a high-bandwidth interface that tightly couples human cognition with robotic execution [63]. By fusing three-dimensional virtual entities with physical or simulated environments, these modalities are revolutionizing the data acquisition and skill learning pipelines for embodied agents. Current VR deployment in robotics demonstrates a stratified landscape: in surgical robotics [64–67] and human-robot interaction (HRI) [68], it enhances operational precision and intuitive collaboration; whereas in motion planning and swarm robotics [69,70], it circumvents the bottlenecks of complex programming and the sensory limitations of micro-robots. Nevertheless, broad adoption remains hindered by ubiquitous engineering challenges, notably acute system complexity, communication latency, and inherent Field of View (FoV) constraints.

To systematically embed human cognitive priors into policy learning, Learning from Demonstration (LfD) has emerged as a pivotal frontier. As illustrated in Fig. 8, an MIT team [71] devised a VR-teleoperated apprenticeship framework combining self-supervised exploration with active learning. During grasping tasks, the system autonomously executes policies, exclusively querying human intervention for high-fidelity demonstrations when confidence deteriorates or consecutive failures occur. Integrated with 3D CNN-based grasp prediction, this active querying mechanism drastically curtails the dependency on continuous human oversight, facilitating scalable data acquisition via a centralized multi-robot monitoring interface. Furthermore, to overcome the interactive bottlenecks of traditional simulators, the iGibson 2.0 platform introduces multidimensional enhancements. It incorporates extended, visually rendered physical states (e.g., temperature and humidity) and employs logical predicate deduction to semanticize environments and automate task generation. Crucially, its native VR interface facilitates direct human demonstration of long-horizon tasks. This synergy of semantic physics and VR-guided teleoperation significantly enriches the dimensionality of imitation learning datasets, thereby augmenting the agent's proficiency in executing logic-driven, multi-step domestic tasks.

While Virtual Reality (VR) is maturely deployed across simulated environments, the efficacy of Augmented Reality (AR) for *in-situ* robotic training remains bottlenecked by camera localization precision and display hardware fidelity. Overcoming these limitations necessitates integrating AI-driven Simultaneous Localization and Mapping (SLAM) algorithms to construct Mixed Reality (MR) frameworks with high-fidelity virtual-physical registration. Ultimately, such paradigms will enable robust augmented training and seamless human-robot interaction directly within unstructured physical domains.

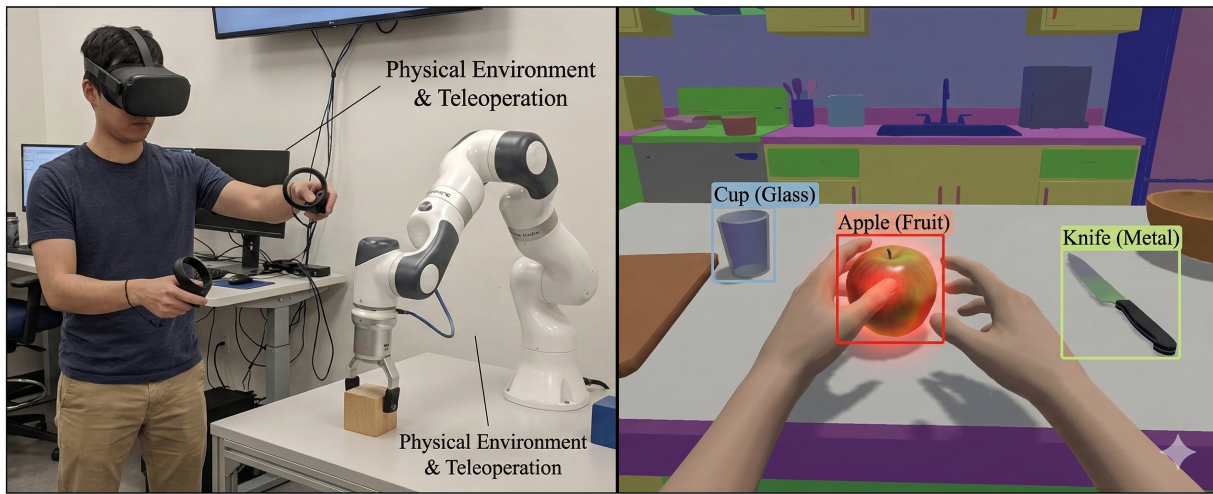


Figure 8: Schematic diagram of VR-assisted robot intelligent training system.

3.2.2 Visual Pre-Training Methodologies

Visual pre-training in embodied AI fundamentally aims to circumvent the severe sample inefficiency intrinsic to tabula rasa learning. By exploiting massive multimodal datasets from diverse domains—including public repositories, simulations, and physical interactions, this paradigm extracts generalizable visual representations through unsupervised [72], self-supervised [73], or weakly supervised [74] frameworks. The core rationale is to embed universal perceptual priors—encompassing object morphology, spatial topologies, and foundational semantics—before exposing the agent to task-specific manipulations. Ultimately, this yields a high-fidelity feature initialization, establishing a robust scaffold for downstream capability transfer in critical operations like object recognition, grasp planning, and spatial navigation.

To efficiently translate generalizable visual representations into executable manipulation skills, recent research has systematically advanced along two primary axes: model transfer strategies and data source diversification. As Table 3 illustrates, Google Brain pioneered the “learning to see before learning to act” paradigm. By repurposing parameters from passive visual tasks via full-architecture transfer, this strategy mitigates the exploration inefficiencies inherent in *de novo* backbone initialization. Furthermore, to embed dynamic human behavioral priors, the R3M model [75] shifts the pre-training corpus from static imagery to Ego4D egocentric videos. While temporal contrastive learning enables this encoder to extract sequential features and boost few-shot performance by over 20%, its overall utility remains bottlenecked by single-frame representational limits and a lack of deep integration with downstream Reinforcement Learning (RL) architectures.

Table 3: Comparison table of core robot vision pre-training models.

Model/Team	Core Strategy & Innovation	Key Technical Mechanism	Performance & Advantages	Main Limitations
Google Brain Team	Full-architecture transfer strategy	Full parameter re-use: Fully transferring both the backbone and prediction heads prevents the training inefficiency of partial transfer.	Improved sample efficiency: Success rate in simulated grasping increased from 71% to 86%.	Lacks utilization of dynamic temporal information; based solely on static visual features.

(Continued)

Table 3 (continued)

Model/Team	Core Strategy & Innovation	Key Technical Mechanism	Performance & Advantages	Main Limitations
R3M [75]	Incorporation of human behavior	Temporal Contrastive Learning: Ex-tracts temporal and behavioral features from video streams to serve as a universal encoder.	Strong few-shot learning: Average success rate across 12 simulated tasks improved by over 20% compared to learning from scratch.	Provides only single-frame representation capability; lacks in-depth exploration of reinforcement learning.
RT-2 [76]	Multimodal end-to-end closed loop	Action Tokenization: Encodes actions into text to-kens and performs co-fine-tuning based on Vision-Language Models (VLMs).	Possesses capabilities for complex instruction parsing, symbolic reasoning, and Chain-of-Thought (CoT) planning.	High computational costs; physical skills are limited by the coverage of the training data.

The emergence of Large Multimodal Models (LMMs), exemplified by Google DeepMind’s RT-2 [76], has fundamentally extended the frontier of embodied pre-training beyond purely visual domains. Built upon a Vision-Language Model (VLM) backbone, RT-2 employs action tokenization to represent robotic kinematics as discrete tokens, enabling the co-fine-tuning of internet-scale vision-language corpora with robotic demonstration data. This perception-reasoning-action integration endows agents with the capacity for complex semantic instruction following, symbolic reasoning, and Chain-of-Thought (CoT) planning. Despite current challenges—such as prohibitive computational overhead and limitations in fine-grained physical skill generalization—RT-2’s manifestation of high-level reasoning marks a pivotal shift: transitioning embodied pre-training from task-specific policy learning toward open-vocabulary and cross-domain generalization.

4 Task Planning Strategies for Embodied Agents in Complex Scenarios

Modern applications of embodied agents are increasingly challenged by unstructured environments and dynamic interactions, where traditional trajectory generation no longer suffices. Consequently, the research focus has pivoted from low-level execution toward cognitive-level in-tent comprehension and integrated Task and Motion Planning (TAMP). This evolution bifurcates into two critical imperatives: first, in human-robot collaboration, agents must transition from passive obstacle avoidance to reactive planning that enables proactive synergy; second, in long-horizon tasks, the core challenge lies in bridging the gap between discrete symbolic logic and continuous geometric motion—a prerequisite for achieving high-level autonomy in embodied intelligence.

4.1 Human-Robot Collaboration and Reactive Planning

As depicted in Fig. 9, reactive planning in Human-Robot Collaboration (HRC) is fundamentally governed by the dual imperatives of safety and efficiency—a mandate that dictates both the algorithmic architecture and its functional adaptation to dynamic contexts. From a technical perspective, this paradigm integrates multi-modal sensor fusion—incorporating vision [77,78], haptics [79,80], and LiDAR [81]—with lightweight inference engines to facilitate the real-time decoding of human intent. Moving beyond rudimentary “reactive avoidance,” this framework enables proactive adaptation via localized environmental

pre-modeling. Its core strength lies in bypassing computationally expensive global replanning, thereby ensuring millisecond-level latency in response to stochastic environmental contingencies.

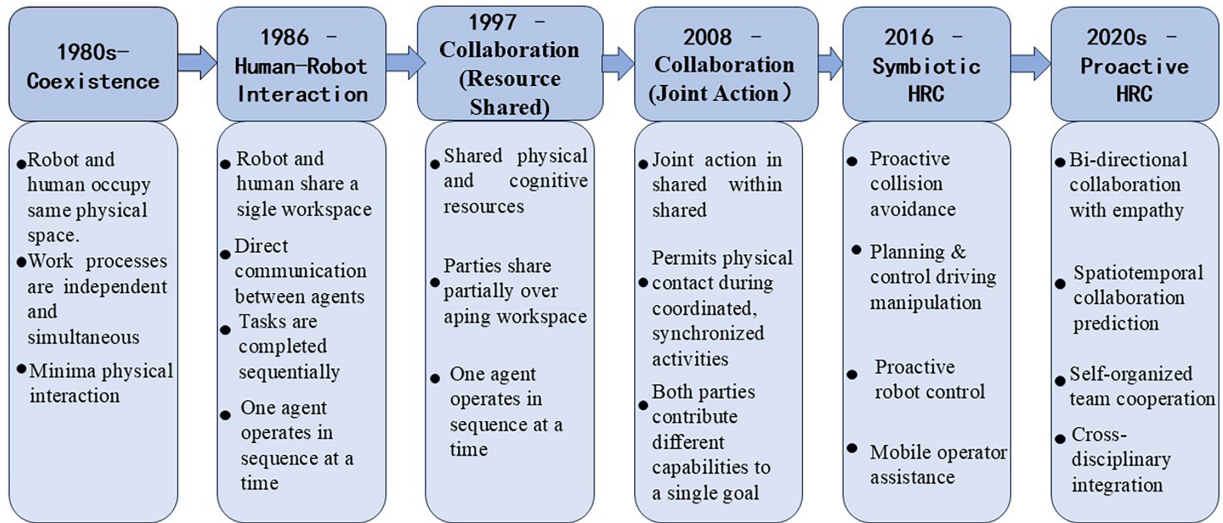


Figure 9: Evolutionary path of proactive human-robot collaboration.

Transitioning from collaborative paradigms to algorithmic execution, Li et al. [82] introduced the Proactive Human-Robot Collaboration (PHRC) model for cognitive manufacturing, specifically addressing the demands of large-scale personalized production. This paradigm leverages digital twins, Augmented Reality (AR), and spatiotemporal trajectory forecasting to evolve the human-robot relationship from mere coexistence toward symbiotic collaboration, thereby mitigating behavioral stochasticity and enhancing system flexibility. To operationalize this proactive ethos within trajectory synthesis, Flowers et al. [83] developed the Spatio-Temporal Avoidance Planning (STAP) framework. By transcending the limitations of static obstacle avoidance, STAP integrates spatiotemporal occupancy mapping with dynamic velocity modulation to enable the precise anticipation of human motion. While computational latency remains a hurdle for real-time deployment, the framework underscores the critical necessity of spatiotemporal prediction in optimizing collaborative fluency.

Regarding specialized process-level applications, Oshin et al. [84] developed a sophisticated collaborative framework for composite material lay-up. The system innovatively synthesizes generative and discriminative Long Short-Term Memory (LSTM) models to accurately forecast human actions, subsequently casting task planning as a Partially Observable Markov Decision Process (POMDP). This mathematical formulation enables dual-arm robots to derive optimal execution policies under perceptual uncertainty, achieving a highly efficient coupling between robotic automation and manual labor in complex industrial workflows.

While existing reactive planning strategies have demonstrated preliminary success in dynamic-static obstacle avoidance, they remain constrained by the inherent trade-off between high-concurrency multimodal processing and real-time decision-making throughput. Despite the influx of spatiotemporal forecasting and deep learning, reconciling high-fidelity prediction with reduced algorithmic complexity remains a critical hurdle for industrial-scale deployment. Current methodologies predominantly stagnate at the geometric level of trajectory prediction, lacking a profound grasp of deep semantic intent. Future research must prioritize the development of light-weight edge-computing frameworks and explore the integration of Large Language Models (LLMs) to facilitate semantic reasoning. This will facilitate a necessary paradigm

shift in embodied AI—moving from action-level anticipation to task-level cognition—to achieve a robust fusion of safety and efficiency in unstructured environments.

4.2 Path Planning for Long-Horizon and Complicated Tasks

Task and Motion Planning (TAMP) acts as a pivotal nexus bridging high-level logical reasoning with low-level physical actuation, thereby enabling embodied agents to achieve sophisticated autonomy. Its core utility lies in the systematic deconstruction of abstract goals into collision-free, executable action sequences within complex environments. Facilitated by the emergence of multi-task learning [85], TAMP methodologies are undergoing a paradigm shift—transitioning from conventional hierarchical decoupling toward a unified integration of symbolic logic and continuous geometry. This evolution propels robotic capabilities beyond rigid automated execution toward adaptive, intelligent decision-making.

To tackle the tight coupling between discrete logical reasoning and continuous motion execution, algorithmic approaches have transitioned from conventional decoupled pipelines to holistic joint optimization strategies. Notably, frameworks like the Graph of Convex Sets (GCS) and Hierarchical Temporal Logic (HTL) [86,87] formulate multi-robot allocation, symbolic reasoning, and trajectory synthesis as a unified shortest-path problem on a product graph. By mapping heterogeneous constraints into a tractable mathematical space, this integrated approach ensures theoretical completeness and global optimality.

Complementing algorithmic optimizations, the integration of Large Language Models (LLMs) catalyzes a cognitive paradigm shift in TAMP, furnishing embodied systems with semantic comprehension and closed-loop reasoning. While traditional TAMP is fundamentally bottlenecked by rigid, hand-crafted heuristics and a deficit in physical common sense, state-of-the-art frameworks like LLM [88] repurpose LLMs as potent heuristic engines. Beyond merely synthesizing high-level task skeletons, these architectures exhibit critical Motion Failure Reasoning (MFR), directly translating geometric collision feedback into autonomous, semantic-level replanning. Concurrently, deep learning-driven planning domain inference [89] automates the extraction of environmental causal models from human demonstrations, effectively decoupling system performance from expert-engineered priors.

To facilitate a structured comparison of the representative TAMP methodologies surveyed above, Table 4 synthesizes their core technical mechanisms, standard performance metrics, principal strengths, and key limitations.

Table 4: Comparison of key mechanisms and performance of mainstream TAMP methods.

Technical Paradigm	Core Mechanism	Standardized Metrics	Advantages	Limitations
Graphs of Convex Sets [83]	Transforms trajectory generation into a shortest-path problem over a sequence of convex sets.	Completeness: High Efficiency: Medium Scalability: Medium. Adaptability: Low	Jointly optimizes discrete paths and continuous trajectories; effectively handles system dynamics.	Highly dependent on the quality of the convex decomposition of the free space.
Product Graphs and HTL [84]	Graph search combining task automata with spatial convex decomposition.	Completeness: High Efficiency: Low Scalability: Low Adaptability: Low	Theoretically complete; guarantees satisfaction of temporal logic; suitable for multi-robot systems.	State space grows exponentially with the number of robots.

(Continued)

Table 4 (continued)

Technical Paradigm	Core Mechanism	Standardized Metrics	Advantages	Limitations
LLM-driven TAMP [85]	Large Language Models (LLMs) act as heuristic generators and failure reasoners.	Completeness: Low Efficiency: Medium Scalability: High Adaptability: High	Possesses common-sense reasoning; handles undefined objects.	Non-deterministic output, lacks guarantees for geometric feasibility.
Partially Instantiated Planning [87]	Alternating execution between planning and online behaviors.	Completeness: Low Efficiency: High Scalability: Medium Adaptability: Medium	Exceptional robustness; adapts well to sensing noise and uncertainty.	Theoretical completeness is difficult to prove; relies heavily on the quality of the behavior library.
Adaptive HRC Planning [88]	Human intent inference and dynamic scheduling based on POMDP.	Completeness: Medium Efficiency: Low Scalability: Low Adaptability: Medium	Enhances fluency and trust in human-robot collaboration (HRC).	Requires an accurate probabilistic model of human behavior.

To counter inevitable perceptual noise and environmental dynamics, Partially Grounded Planning (PGP) introduces a highly resilient paradigm [90]. By deferring concrete geometric computations until execution—where abstract plans are directly reconciled with real-time sensory data—PGP significantly enhances system robustness. Concurrently, psychology informed TAMP algorithms now enable real-time human intent inference and dynamic role modulation [91]. Together, these advancements propel embodied agents from sequestered automation cells into unstructured, symbiotic environments.

In applied domains, recent literature highlights both critical milestones and persistent bottlenecks. For high-precision assembly, Huang et al. [92] developed Choreo, leveraging hierarchical and semi-constrained Cartesian planning to execute spatial truss assembly. While computationally efficient, full autonomy remains hindered by its reliance on manual task decomposition and the absence of automated backtracking. Addressing dynamic environments, Bernardo et al. [93] proposed a closed-loop semantic reasoning framework bridging perceptual detection with collision-free trajectory generation. Despite robust logical capabilities, the system requires further refinement to manage severe state uncertainty and integrate mobile bases. Finally, in large-scale multi-agent coordination, Yan and Julius [94] introduced a decentralized branch-and-bound algorithm utilizing mixed-integer programming to balance computational loads. Although this approach substantially accelerates solving speeds, striking an optimal trade-off between communication overhead and the optimality loss inherent in approximate solvers remains a pivotal hurdle for real-world deployment.

5 Cooperative Control in Embodied Agents

Decentralized multi-agent coordination has emerged as a prominent approach for deploying embodied robots in dynamic, large-scale environments, effectively mitigating the computational bottlenecks often associated with purely centralized architectures. While conventional centralized control relying on a global planner—provides strong guarantees of global optimality, its scalability is frequently constrained by exponential computational complexity and vulnerability to single points of failure as the agent population expands. However, rather than a complete paradigm shift toward pure decentralization, contemporary

research and practical deployments increasingly favor hybrid or hierarchical frameworks. These architectures synthesize the strengths of both paradigms: they leverage centralized high-level orchestration for global coherence and task allocation, while delegating low-level reactive execution to decentralized, autonomous agents. This balanced topology ensures robust real-time performance without sacrificing global optimization.

Within these hierarchical structures, the decentralized execution layer decouples real-time functionality from the global node, endowing individual agents with localized autonomous decision-making. By leveraging local perception, constrained peer-to-peer communication, and independent reasoning, agents can rapidly execute dynamic trajectory synthesis and localized conflict resolution. While purely decentralized interactions may struggle to guarantee strict global optimality, they excel in generating resilient, emergent responses to environmental perturbations. Consequently, the inherent scalability and adaptability of decentralized mechanisms render them an indispensable component—rather than a standalone definitive solution—for navigating highly uncertain, large-scale mission spaces.

As depicted in Fig. 10, the paradigm of decentralized multi-agent coordination has progressively advanced from low-level perceptual interactions to large-scale swarm planning and high-level decision-making. Addressing initial bottlenecks in intent inference and bandwidth, Wu et al. [95] introduced “spatial intent.” By transmitting 2D key points via Double Deep Q-Networks (DDQN), their approach aligns action intentions and ensures robust sim-to-real transfer under severe communication constraints. Scaling up to complex Multi-Agent Path Finding (MAPF), Sartoretti et al. [96] proposed the PRIMAL framework. Hybridizing reinforcement and imitation learning, PRIMAL achieves millisecond-level, collision-free planning for up to a thousand agents without explicit communication, effectively shattering the computational limits of centralized processing. Further advancing integrated task and path optimization, Chen et al. [97] developed DTPP. By leveraging the Max-Sum algorithm and Mixed Observability Markov Decision Processes (MOMDPs) to embed cost-benefit analysis into dynamic task allocation, DTPP elevates multi-agent synergy from rudimentary collision avoidance to the maximization of net team utility.

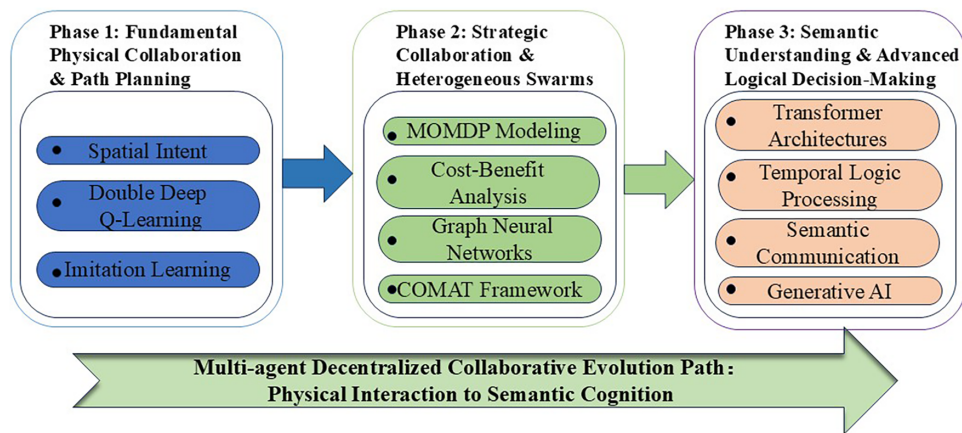


Figure 10: Evolutionary roadmap of multi-agent decentralized collaboration technology.

Building upon the foundational concept of spatial intent for multi-agent coordination, integrating Graph Neural Networks (GNNs) and Transformers has propelled multi-agent systems from rudimentary collision avoidance to profound semantic comprehension. Frameworks like DGNN-GA [98] and advanced GNNs [99] effectively resolve robustness bottlenecks in large-scale dynamic allocation under volatile communication, while COMAT [100] achieves flexible combinatorial generalization across heterogeneous

teams. Further augmentations equip these architectures to manage complex temporal logic and resolve traffic deadlocks [101,102], alongside leveraging generative AI and semantic communication to sustain coordination under ultra-low bandwidths [103]. Collectively, these breakthroughs yield decentralized, intelligent swarms equipped to navigate extreme environments, interpret natural language directives, and strictly adhere to logical constraints.

Decentralized multi-agent coordination has evolved from fundamental collision avoidance to advanced semantic cognition. Looking ahead, the fusion of embodied foundation models with world models offers a pathway to liberate these systems from task-specific constraints. However, realizing generalized swarm intelligence in open-world environments requires overcoming specific technical bottlenecks. Future research must prioritize:

- (1) robust uncertainty handling to manage noisy multi-modal perception and stochastic environmental perturbations.
- (2) efficient real-time inference to resolve the latency inherent in deploying large-parameter models across edge devices for high-frequency control.
- (3) precise semantic grounding to seamlessly translate high-level natural language reasoning into physically executable, low-level kinematic actions.

Ultimately, resolving these concrete challenges to achieve deep alignment between swarm dynamics and human intent will unlock the transformative potential of embodied agents in large-scale ecosystems.

6 Discussion and Future Perspectives

Modeling manipulation skills for embodied agents has emerged as a frontier domain at the nexus of robotics, artificial intelligence, and control theory. Recent literature has driven substantial progress across four core technical pillars: motion control optimization, intelligent training methods, complex task planning, and multi-agent collaboration. However, current models still fall critically short of the dynamic, high-fidelity physical interaction demanded by unstructured real-world environments. Synthesizing the established state-of-the-art, this section critically examines prevailing bottlenecks and outlines future trajectories to advance the field.

6.1 Existing Limitations

- (1) Inadequate Mapping Between Complex Physical Interactions and Environmental Perception.

Despite breakthroughs in rigid-body simulation, the Sim-to-Real gap remains a formidable obstacle for scenarios involving complex friction, soft-body deformation, and fluid interactions. The complexity of the physical world is fundamentally irreducible; simulators constrained by idealized models often fail to capture the micro-characteristics of unstructured environments. For instance, the stochastic twisting of flexible cables under gravitational and contact forces is notoriously difficult to replicate in simulation. Furthermore, a critical disconnect persists between high-level semantic planning and low-level geometric constraints. While LLM-driven planners exhibit sophisticated reasoning, their lack of intuitive physical grounding frequently results in “physically infeasible” long-horizon plans [104] that overlook kinematic limits or collision volumes. Crucially, vulnerabilities caused by “LLM hallucinations” where models confidently generate logically plausible but physically incorrect actions—introduce systemic safety risks during physical human-robot interactions. When ungrounded semantic plans are directly translated into execution, they severely compromise the reliability of the system in dynamic, collaborative spaces.

(2) Performance Trade-offs Between Generalization and Precision in Embodied Foundation Models.

Current research trajectories reveal a critical paradox: the pursuit of model universality often compromises task-specific accuracy and execution reliability. While foundation models, such as Open X-Embodiment [105], have significantly enhanced generalization across unseen scenarios, this versatility typically comes at the cost of diminished precision in specialized tasks. Such large-scale models are prone to learning the “mean distribution” of manipulation data, thereby marginalizing the stringent accuracy requirements of niche domains. Consequently, they frequently underperform in tasks demanding sub-millimeter force-feedback control or high-fidelity motion precision [106]. In the context of real-world deployment, this broad statistical generalization fails to satisfy the deterministic rigor required for safe, fine-grained manipulation, leading to unpredictable failure modes.

(3) Computational Overhead and Latency Constraints in Engineering Deployment.

The widespread application of embodied AI is currently stifled by stringent physical constraints and real-time processing bottlenecks inherent in advanced algorithms. Specifically, the massive computational footprint of Transformer architectures and high-dimensional dynamic forecasting often clashes with the computational overhead limits on edge hardware. In dynamic scenarios necessitating millisecond-level responsiveness, the inferential latency of large-scale models fails to meet the strict latency requirements for real-time reactive control, frequently resulting in delayed actuation [107] or hazardous collision events. Furthermore, deploying these compute-intensive frameworks on battery-powered agents leads to excessive power dissipation, drastically curtailing operational endurance and system reliability. Such limitations collectively undermine the commercial viability and safe, large-scale deployment of embodied intelligent systems in unconstrained real-world environments.

6.2 Future Perspectives

(1) Physics-Informed Cross-Modal World Models:

To transcend current perceptual limitations, future embodied agents must evolve toward unified, cross-modal world models that synergistically integrate visual, tactile, auditory, and proprioceptive streams. By leveraging Physics-Informed Neural Networks (PINNs) and Koopman operator theory, physical conservation laws can be embedded directly into neural network loss functions, enabling agents to internalize and extrapolate fundamental physical principles. This structural fusion facilitates zero-shot causal reasoning and predictive modeling in novel or partially observable environments, thereby fundamentally fortifying algorithmic robustness against real-world stochasticity.

(2) Generative Simulation-Driven Pre-training and Lifelong Learning Loops:

To address the persistent challenges of data scarcity and Sim-to-Real distributional shifts, future technical architectures will pivot toward a hybrid paradigm that couples generative simulation pre-training with real-world adaptive fine-tuning. Leveraging generative AI, high-fidelity simulators can autonomously synthesize infinite virtual environments, specifically targeting the “long-tail” distribution to provide exhaustive datasets. More importantly, embodied agents must integrate lifelong learning mechanisms to sustain a continuous data-collection cycle during operational deployment. By extracting insights from task failures and iteratively updating policy networks, agents can achieve self-directed capability evolution and consistent performance enhancement in unconstrained environments.

(3) Cloud-Edge-End Synergy and Human-Centric Value Alignment:

Future system architectures will likely adopt 5G/6G-enabled semantic communication to facilitate seamless cloud-edge-end orchestration, effectively reconciling the inherent conflict between high computational intensity and real-time execution. Ultimately, the development of embodied AI must pivot back to the human element. Future research should transcend objective metrics, such as task completion rates, to prioritize deep human-robot value alignment. This shift necessitates that agents not only infer implicit human intentions but also treat human safety and comfort as invariant constraints during physical interactions, thereby fostering a harmonious co-existence within large-scale, complex societal ecosystems.

7 Conclusion

This paper provides a comprehensive review of the development and implementation of manipulation skill models for embodied intelligent robots. By systematically evaluating the advancements in motion control, intelligent training, task planning, and multi-agent collaboration, we have highlighted the transformative potential of blending deep learning with control theory. Although persistent challenges remain regarding the Sim-to-Real gap, precision-generalization trade-offs, and computational latency on edge hardware, emerging paradigms such as physics-informed world models, generative pre-training, and cloud-edge synergy offer promising pathways forward. Ultimately, bridging these gaps will not only enhance the operational autonomy and robustness of embodied agents in unstructured environments but also ensure their seamless and safe integration into human-centric societal ecosystems.

Acknowledgement: The authors would like to express their sincere gratitude to the laboratory staff and technical support personnel for their indispensable assistance in data acquisition and the provision of computing infrastructure.

Funding Statement: This review was funded by Youth Fund of the National Natural Science Foundation of China, grant number 62406032; Beijing Natural Science Foundation, grant number 4242036; The 2025 Open Fund of the State Key Laboratory of Space Intelligent Control Technology grant number HTKJ2025KL502016.

Author Contributions: Conceptualization: Lianpeng Li and Zhoujun Ruan; methodology: Lianpeng Li and Zhoujun Ruan; software: Zhichuang Wang; validation: Lianpeng Li, Zhoujun Ruan and Haibo Zhang; formal analysis: Lianpeng Li and Zhoujun Ruan; investigation: Zhoujun Ruan and Hang Zhong; resources: Lianpeng Li; writing—original draft preparation: Zhoujun Ruan; writing—review and editing, Lianpeng Li, Zhoujun Ruan, Mingyang Li and Chunpeng Kang; supervision: Lianpeng Li and Chunpeng Kang; funding acquisition: Lianpeng Li. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Supplementary Materials: The supplementary material is available online at <https://www.techscience.com/doi/10.32604/cmc.2026.084092/s1>.

References

1. Feng Z, Xue R, Yuan L, Yu Y, Ding N, Liu M, et al. Multi-agent embodied AI: advances and future directions. *Sci China Inf Sci*. 2026;69(5):151202. doi:10.1007/s11432-025-4820-4.
2. Feng T, Wang X, Jiang YG, Zhu W. Embodied AI: from LLMs to world models. Preprint. 2025. doi:10.36227/techrxiv.175977432.27129012/v1.

3. Dutta A, Burdet E, Kaboli M. ViTract: robust object shape perception via active visuo-tactile interaction. *IEEE Robot Autom Lett.* 2024;9(12):11250–7. doi:10.1109/lra.2024.3483037.
4. Xu Z, Wu K, Wen J, Li J, Liu N, Che Z, et al. A survey on robotics with foundation models: toward embodied AI. *arXiv:2402.02385.* 2024.
5. Nilsson NJ. *Shakey the robot.* Menlo Park, CA, USA: SRI International; 1984.
6. Sakagami Y, Watanabe R, Aoyama C, Matsunaga S, Higaki N, Fujimura K. The intelligent ASIMO: system overview and integration. In: *Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2002 Sep 30–Oct 4; Lausanne, Switzerland.* p. 2478–83.
7. Tsagarakis NG, Metta G, Sandini G, Vernon D, Beira R, Becchi F, et al. iCub: the design and realization of an open humanoid platform for cognitive and neuroscience research. *Adv Robot.* 2007;21(10):1151–75. doi:10.1163/156855307781389419.
8. Xu J, Sun Q, Han QL, Tang Y. When embodied AI meets industry 5.0: human-centered smart manufacturing. *IEEE CAA J Autom Sin.* 2025;12(3):485–501. doi:10.1109/jas.2025.125327.
9. Lisondra M, Benhabib B, Nejat G. Embodied AI with foundation models for mobile service robots: a systematic review. *arXiv:2505.20503.* 2025.
10. Rao H, Bi X, Wang Z, Chen Q. Research on special operation robot technology for aircraft inlet. In: *Proceedings of the 2024 6th International Conference on Internet of Things, Automation and Artificial Intelligence (IoTAAI); 2024 Jul 26–8; Guangzhou, China.* p. 183–9. doi:10.1109/IoTAAI62601.2024.10692670.
11. He Z, Fang H, Chen J, Fang HS, Lu C. FoAR: force-aware reactive policy for contact-rich robotic manipulation. *IEEE Robot Autom Lett.* 2025;10(6):5625–32. doi:10.1109/lra.2025.3560871.
12. Li G, Liang X, Gao Y, Su T, Liu Z, Hou ZG. A linkage-driven underactuated robotic hand for adaptive grasping and in-hand manipulation. *IEEE Trans Automat Sci Eng.* 2024;21(3):3039–51. doi:10.1109/tase.2023.3273721.
13. Zeng T, Mohammad A, Madrigal AG, Axinte D, Keedwell M. A robust human–robot collaborative control approach based on model predictive control. *IEEE Trans Ind Electron.* 2024;71(7):7360–9. doi:10.1109/tie.2023.3299046.
14. Asker MA, Gaeid KS, Tawfeeq NN, Zain HK, Kauther AI, Abdullah Q, et al. Design and analysis of robot PID controller using digital signal processing techniques. *Int J Eng Technol.* 2018;4(37):103–9. doi:10.14419/ijet.v7i4.37.23625.
15. Huang X, Guo W, Liu S, Li Y, Qiu Y, Fang H, et al. Flexible mechanical metamaterials enabled electronic skin for real-time detection of unstable grasping in robotic manipulation. *Adv Funct Mater.* 2022;32(23):2109109. doi:10.1002/adfm.202109109.
16. Hyun NP, Vela PA, Verriest EI. Collision free and permutation invariant formation control using the root locus principle. In: *Proceedings of the 2016 American Control Conference (ACC); 2016 Jul 6–8; Boston, MA, USA.* p. 2572–7. doi:10.1109/acc.2016.7525304.
17. Guffanti D, Brunete A, Hernando Gutierrez M. Human gait analysis using non-invasive methods with a ROS-based mobile robotic platform. In: *Proceedings of the New Trends in Medical and Service Robotics; 2020 Jul 8–10; Basel, Switzerland.* p. 309–17. doi:10.1007/978-3-030-58104-6_35.
18. Cai XL, Yang WA, You YP. Multi-mode frequency response prediction of milling robot based on feature transferring with small sample sets. *J Vibroeng.* 2025;27(6):1135–58. doi:10.21595/jve.2025.25099.
19. Karthik R, Menaka R, Kishore P, Aswin R, Vikram C. Dual mode PID controller for path planning of encoder less mobile robots in warehouse environment. *IEEE Access.* 2024;12(10):21634–46. doi:10.1109/ACCESS.2024.3363898.
20. Wu YH, Yu ZC, Li CY, He MJ, Hua B, Chen ZM. Reinforcement learning in dual-arm trajectory planning for a free-floating space robot. *Aerosp Sci Technol.* 2020;98(1):105657. doi:10.1016/j.ast.2019.105657.
21. Srivastava N, Talbott W, Lopez MB, Zhai S, Susskind J. Robust robotic control from pixels using contrastive recurrent state-space models. *arXiv:2112.01163.* 2021.
22. Wang L, Wang G, Jia S, Turner A, Ratchev S. Imitation learning for coordinated human-robot collaboration based on hidden state-space models. *Robot Comput Integr Manuf.* 2022;76(1–4):102310. doi:10.1016/j.rcim.2021.102310.

23. Teng S, Mueller MW, Sreenath K. Legged robot state estimation in slippery environments using invariant extended Kalman filter with velocity update. In: Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA); 2021 May 30–Jun 5; Xi'an, China. p. 3104–10. doi:10.1109/icra48506.2021.9561313.
24. Xu X, Pang F, Ran Y, Bai Y, Zhang L, Tan Z, et al. An indoor mobile robot positioning algorithm based on adaptive federated Kalman filter. *IEEE Sens J*. 2021;21(20):23098–107. doi:10.1109/jsen.2021.3106301.
25. Abadía I, Naveros F, Ros E, Carrillo RR, Luque NR. A cerebellar-based solution to the nondeterministic time delay problem in robotic control. *Sci Robot*. 2021;6(58):eabf2756. doi:10.1126/scirobotics.abf2756.
26. Sakib M S, Sun Y. STAR: a foundation model-driven framework for robust task planning and failure recovery in robotic systems. *Int J Artif Intell Robot Res*. 2025;2:2550007.
27. Nagatani K, Yamasaki A, Yoshida K, Yoshida T, Koyanagi E. Semi-autonomous traversal on uneven terrain for a tracked vehicle using autonomous control of active flippers. In: Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2008 Sep 22–26; Nice, France. p. 2667–72. doi:10.1109/IROS.2008.4650643.
28. Ohno K, Morimura S, Tadokoro S, Koyanagi E, Yoshida T. Semi-autonomous control system of rescue crawler robot having flippers for getting over unknown-steps. In: Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2007 Oct 29–Nov 2; San Diego, CA, USA. p. 3012–8. doi:10.1109/IROS.2007.4399271.
29. Okada Y, Nagatani K, Yoshida K, Tadokoro S, Yoshida T, Koyanagi E. Shared autonomy system for tracked vehicles on rough terrain based on continuous three-dimensional terrain scanning. *J Field Robot*. 2011;28(6):875–93. doi:10.1002/rob.20416.
30. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature*. 2015;518(7540):529–33. doi:10.1038/nature14236.
31. Mitriakov A, Papadakis P, Mai Nguyen S, Garlatti S. Staircase negotiation learning for articulated tracked robots with varying degrees of freedom. In: Proceedings of the 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR); 2020 Nov 4–6; Abu Dhabi, United Arab Emirates. p. 394–400. doi:10.1109/ssrr50563.2020.9292594.
32. Mitriakov A, Papadakis P, Mai Nguyen S, Garlatti S. Staircase traversal via reinforcement learning for active reconfiguration of assistive robots. In: Proceedings of the 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE); 2020 Jul 19–24; Glasgow, UK. p. 1–8. doi:10.1109/fuzz48607.2020.9177581.
33. Zimmermann K, Zuzanek P, Reinstein M, Hlavac V. Adaptive traversability of unknown complex terrain with obstacles for mobile robots. In: Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA); 2014 May 31–Jun 7; Hong Kong, China. p. 5177–82. doi:10.1109/ICRA.2014.6907619.
34. Pan H, Chen X, Ren J, Chen B, Huang K, Zhang H, et al. Deep reinforcement learning for flipper control of tracked robots in urban rescuing environments. *Remote Sens*. 2023;15(18):4616. doi:10.3390/rs15184616.
35. Grandia R, Farshidian F, Knoop E, Schumacher C, Hutter M, Bächer M. DOC: differentiable optimal control for retargeting motions onto legged robots. *ACM Trans Graph*. 2023;42(4):1–14. doi:10.1145/3592454.
36. Bjelonic M, Sankar PK, Bellicoso CD, Vallery H, Hutter M. Rolling in the deep-hybrid locomotion for wheeled-legged robots using online trajectory optimization. *IEEE Robot Autom Lett*. 2020;5(2):3626–33. doi:10.1109/lra.2020.2979661.
37. Carius J, Ranftl R, Koltun V, Hutter M. Trajectory optimization for legged robots with slipping motions. *IEEE Robot Autom Lett*. 2019;4(3):3013–20. doi:10.1109/lra.2019.2923967.
38. Ren AZ, Lidard J, Ankile LL, Simeonov A, Agrawal P, Majumdar A, et al. Diffusion policy policy optimization. *arXiv:2409.00588*. 2024.
39. Liu Y, Bai Y, Lin L. Technological system for embodied intelligence towards efficient integration and coordination of human, robot, and physical world. *Robot*. 2025;47(4):559–80. (In Chinese). doi:10.13973/j.cnki.robot.250191.
40. NVIDIA. GR00T N1: an open foundation model for generalist humanoid robots. *arXiv:2503.14734*. 2025.
41. Zhao Q, Lu Y, Kim MJ, Fu Z, Zhang Z, Wu Y, et al. CoT-VLA: visual chain-of-thought reasoning for vision-language-action models. In: Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2025 Jun 10–7; Nashville, TN, USA. p. 1702–13. doi:10.1109/cvpr52734.2025.00166.

42. Yuan K, Sajid N, Friston K, Li Z. Hierarchical generative modelling for autonomous robots. *Nat Mach Intell.* 2023;5(12):1402–14. doi:10.1038/s42256-023-00752-z.
43. Katayama S, Murooka M, Tazaki Y. Model predictive control of legged and humanoid robots: models and algorithms. *Adv Robot.* 2023;37(5):298–315. doi:10.1080/01691864.2023.2168134.
44. Wang Y, Yang M, Ding Z, Zhang Y, Zeng W, Xu X, et al. From experts to a generalist: toward general whole-body control for humanoid robots. *arXiv:2506.12779.* 2025.
45. Zhang Q, Han G, Sun J, Zhao W, Cao J, Wang J, et al. LiPS: large-scale humanoid robot reinforcement learning with parallel-series structures. *arXiv:2503.08349.* 2025.
46. Chen X, Hu J, Jin C, Li L, Wang L. Understanding domain randomization for sim-to-real transfer. *arXiv:2110.03239.* 2021.
47. Kiumarsi B, Vamvoudakis KG, Modares H, Lewis FL. Optimal and autonomous control using reinforcement learning: a survey. *IEEE Trans Neural Netw Learning Syst.* 2018;29(6):2042–62. doi:10.1109/tnnls.2017.2773458.
48. Chen F, Szenher P, Huang Y, Wang J, Shan T, Bai S, et al. Zero-shot reinforcement learning on graphs for autonomous exploration under uncertainty. In: *Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA); 2021 May 30–Jun 5; Xi'an, China.* p. 5193–9. doi:10.1109/icra48506.2021.9561917.
49. Kober J, Bagnell JA, Peters J. Reinforcement learning in robotics: a survey. *Int J Robot Res.* 2013;32(11):1238–74. doi:10.1177/0278364913495721.
50. Wang J, Ma Y, Zhang L, Gao RX, Wu D. Deep learning for smart manufacturing: methods and applications. *J Manuf Syst.* 2018;48(2):144–56. doi:10.1016/j.jmsy.2018.01.003.
51. Zhao W, Queralta JP, Westerlund T. Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In: *Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI); 2020 Dec 1–4; Canberra, Australia.* p. 737–44. doi:10.1109/ssci47803.2020.9308468.
52. Shridhar M, Thomason J, Gordon D, Bisk Y, Han W, Mottaghi R, et al. ALFRED: a benchmark for interpreting grounded instructions for everyday tasks. In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–9; Seattle, WA, USA.* p. 10737–46. doi:10.1109/cvpr42600.2020.01075.
53. Ibarz J, Tan J, Finn C, Kalakrishnan M, Pastor P, Levine S. How to train your robot with deep reinforcement learning: lessons we have learned. *Int J Robot Res.* 2021;40(4–5):698–721. doi:10.1177/0278364920987859.
54. Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA. Deep reinforcement learning: a brief survey. *IEEE Signal Process Mag.* 2017;34(6):26–38. doi:10.1109/msp.2017.2743240.
55. Van Seijen H, Fatemi M, Romoff J, Laroche R, Barnes T, Tsang J. Hybrid reward architecture for reinforcement learning. *Adv Neural Inf Process Syst.* 2017;30:1–11.
56. Huang C, Wang G, Zhou Z, Zhang R, Lin L. Reward-adaptive reinforcement learning: dynamic policy gradient optimization for bipedal locomotion. *IEEE Trans Pattern Anal Mach Intell.* 2023;45(6):7686–95. doi:10.1109/TPAMI.2022.3223407.
57. Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. *arXiv:1509.02971.* 2015.
58. Wu J, Liu Y, Man Z, Sun Z, Yang X, Cao X. Data-driven dynamics modeling of a 9-degree-of-freedom rehabilitation robot based on the Koopman operator. In: *Proceedings of the 2025 IEEE International Instrumentation and Measurement Technology Conference (I2MTC); 2025 May 19–22; Chemnitz, Germany.* p. 1–5. doi:10.1109/I2MTC62753.2025.11079056.
59. Haarnoja T, Ben M, Lever G, Huang SH, Tirumala D, Humplik J, et al. Learning agile soccer skills for a bipedal robot with deep reinforcement learning. *Sci Robot.* 2024;9(89):eadi8022. doi:10.1126/scirobotics.adi8022.
60. Zhu W, Raza F, Hayashibe M. Reinforcement learning based hierarchical control for path tracking of a wheeled bipedal robot with sim-to-real framework. In: *Proceedings of the 2022 IEEE/SICE International Symposium on System Integration (SII); 2022 Jan 9–12; Narvik, Norway.* p. 40–6. doi:10.1109/sii52469.2022.9708882.
61. Siekmann J, Godse Y, Fern A, Hurst J. Sim-to-real learning of all common bipedal gaits via periodic reward composition. In: *Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA); 2021 May 30–Jun 5; Xi'an, China.* p. 7309–15. doi:10.1109/icra48506.2021.9561814.

62. Rodriguez D, Behnke S. DeepWalk: omnidirectional bipedal gait by deep reinforcement learning. In: Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA); 2021 May 30–Jun 5; Xi'an, China. p. 3033–9. doi:10.1109/icra48506.2021.9561717.
63. Makhataeva Z, Varol HA. Augmented reality for robotics: a review. *Robotics*. 2020;9(2):1–28. doi:10.3390/robotics9020021.
64. Wang Q, Zhao K, Song G, Zhao Y, Zhao X. Augmented reality-based navigation system for minimally invasive spine surgery. *Robot*. 2023;45(5):546–53. (In Chinese). doi:10.13973/j.cnki.robot.220300.
65. Karimi A, Moalemi S, Pedramfard P, Moalemi A. Merging robotics and virtual reality for brain tumor surgery: a new frontier in surgical innovation. *World Neurosurg*. 2026;205:124715. doi:10.1016/j.wneu.2025.124715.
66. Moglia A, Ferrari V, Morelli L, Ferrari M, Mosca F, Cuschieri A. A systematic review of virtual reality simulators for robot-assisted surgery. *Eur Urol*. 2016;69(6):1065–80. doi:10.1016/j.eururo.2015.09.021.
67. Cruz-Neira C, Fernández M, Portalés C. Virtual reality and games. *Multimodal Technol Interact*. 2018;2(1):8. doi:10.3390/mti2010008.
68. Chen Y, Zhang B, Zhou J, Wang K. Real-time 3D unstructured environment reconstruction utilizing VR and Kinect-based immersive teleoperation for agricultural field robots. *Comput Electron Agric*. 2020;175(3):105579. doi:10.1016/j.compag.2020.105579.
69. Weller R, Schroder C, Teuber J, Dittmann P, Zachmann G. VR-interactions for planning planetary swarm exploration missions in VaMEx-vtb. In: Proceedings of the 2021 IEEE Aerospace Conference (50100); 2021 Mar 6–13; Big Sky, MT, USA. p. 1–11. doi:10.1109/aero50100.2021.9438374.
70. Ichihashi S, Kuroki S, Nishimura M, Kasaura K, Hiraki T, Tanaka K, et al. Swarm body: embodied swarm robots. *Comput Res Repos*. 2024;267:1–19. doi:10.1145/3613904.3642870.
71. Li C, Xia F, Martín-Martín R, Lingelbach M, Srivastava S, Shen B, et al. iGibson 2.0: object-centric simulation for robot learning of everyday household tasks. *arXiv:2108.03272*. 2021.
72. Pinheiro PO, Almahairi A, Benmalek RY, Golemo F, Courville A. Unsupervised learning of dense visual representations. *arXiv:2011.05499*. 2020.
73. Liu X, Zhang F, Hou Z, Mian L, Wang Z, Zhang J, et al. Self-supervised learning: generative or contrastive. *IEEE Trans Knowl Data Eng*. 2023;35(1):857–76. doi:10.1109/TKDE.2021.3090866.
74. Li YF, Guo LZ, Zhou ZH. Towards safe weakly supervised learning. *IEEE Trans Pattern Anal Mach Intell*. 2021;43:334–46. doi:10.1109/tpami.2019.2922396.
75. Nair S, Rajeswaran A, Kumar V, Finn C, Gupta A. R3M: a universal visual representation for robot manipulation. *arXiv:2203.12601*. 2022.
76. Zitkovich B, Yu T, Xu S, Xu P, Xiao T, Xia F, et al. RT-2: vision-language-action models transfer web knowledge to robotic control. *Comput Res Repos*. 2023;229:2165–83.
77. Araujo HL, Agudelo JG, Vidal RC, Uribe JA, Remolina JF, Serpa-Imbett C, et al. Autonomous mobile robot implemented in LEGO EV3 integrated with raspberry pi to use android-based vision control algorithms for human-machine interaction. *Machines*. 2022;10(3):1–20. doi:10.3390/machines10030193.
78. Chen J, Liu Y, Carey SJ, Dudek P. Proximity estimation using vision features computed on sensor. In: Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA); 2020 May 31–Aug 31; Paris, France. p. 2689–95. doi:10.1109/icra40945.2020.9197370.
79. Costanzo M, Stelter S, Natale C, Pirozzi S, Bartels G, Maldonado A, et al. Manipulation planning and control for shelf replenishment. *IEEE Robot Autom Lett*. 2020;5(2):1595–601. doi:10.1109/lra.2020.2969179.
80. Costanzo M, Maria GD, Lettera G, Natale C, Pirozzi S. Motion planning and reactive control algorithms for object manipulation in uncertain conditions. *Robotics*. 2018;7(4):76. doi:10.3390/robotics7040076.
81. Reddy AK, Malviya V, Kala R. Social cues in the autonomous navigation of indoor mobile robots. *Int J Soc Robotics*. 2021;13(6):1335–58. doi:10.1007/s12369-020-00721-1.
82. Li S, Wang R, Zheng P, Wang L. Towards proactive human-robot collaboration: a foreseeable cognitive manufacturing paradigm. *J Manuf Syst*. 2021;60:547–52. doi:10.1016/j.jmsy.2021.07.017.
83. Flowers J, Faroni M, Wiens G, Pedrocchi N. Spatio-temporal avoidance of predicted occupancy in human-robot collaboration. In: Proceedings of the 2023 32nd IEEE International Conference on Robot and Human Interactive

- Communication (RO-MAN); 2023 Aug 28–31; Busan, Republic of Korea. p. 2162–8. doi:10.1109/ro-man57019.2023.10309469.
84. Oshin O, Bernal EA, Nair BM, Ding J, Varma R, Osborne RW, et al. Coupling deep discriminative and generative models for reactive robot planning in human-robot collaboration. In: *Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*; 2019 Oct 6–9; Bari, Italy. p. 1869–74. doi:10.1109/smc.2019.8913974.
 85. Zheng Y, Yao L, Su Y, Zhang Y, Wang Y, Zhao S, et al. A survey of embodied learning for object-centric robotic manipulation. *Mach Intell Res.* 2025;22(4):588–626. doi:10.1007/s11633-025-1542-8.
 86. Wei Z, Luo X, Liu C. Hierarchical temporal logic task and motion planning for multi-robot systems. *arXiv:2504.18899*. 2025.
 87. Zhao Z, Cheng S, Ding Y, Zhou Z, Zhang S, Xu D, et al. A survey of optimization-based task and motion planning: from classical to learning approaches. *IEEE/ASME Trans Mechatron.* 2025;30(4):2799–825. doi:10.1109/tmech.2024.3452509.
 88. Wang S, Han M, Jiao Z, Zhang Z, Wu YN, Zhu SC, et al. LLM3: large language model-based task and motion planning with motion failure reasoning. In: *Proceedings of the 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; 2024 Oct 14–8; Abu Dhabi, United Arab Emirates. p. 12086–92. doi:10.1109/IROS58592.2024.10801328.
 89. Huang J, Tao A, Marco R, Bogdanovic M, Kelly J, Shkurti F. Automated planning domain inference for task and motion planning. In: *Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA)*; 2025 May 19–23; Atlanta, GA, USA. p. 12534–40. doi:10.1109/ICRA55743.2025.11127817.
 90. Pan T, Shome R, Kavraki LE. Task and motion planning for execution in the real. *IEEE Trans Robot.* 2024;40(1):3356–71. doi:10.1109/TRO.2024.3418550.
 91. Noormohammadi-Asl A, Smith SL, Dautenhahn K. To lead or to follow? Adaptive robot task planning in human-robot collaboration. *IEEE Trans Robot.* 2025;41:4215–35. doi:10.1109/tro.2025.3582816.
 92. Huang Y, Garrett CR, Mueller CT. *Automated motion planning for robotic assembly of discrete architectural structures*. Cambridge, MA, USA: Massachusetts Institute of Technology; 2018. p. 1–22.
 93. Bernardo R, Sousa JMC, Gonçalves PJS. A novel framework to improve motion planning of robotic systems through semantic knowledge-based reasoning. *Comput Ind Eng.* 2023;182(20):109345. doi:10.1016/j.cie.2023.109345.
 94. Yan R, Julius A. A decentralized B&B algorithm for motion planning of robot swarms with temporal logic specifications. *IEEE Robot Autom Lett.* 2021;6(4):7389–96. doi:10.1109/LRA.2021.3098059.
 95. Wu J, Sun X, Zeng A, Song S, Rusinkiewicz S, Funkhouser T. Spatial intention maps for multi-agent mobile manipulation. In: *Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA)*; 2021 May 30–Jun 5; Xi'an, China. p. 8749–56. doi:10.1109/ICRA48506.2021.9561359.
 96. Sartoretti G, Kerr J, Shi Y, Wagner G, Kumar TKS, Koenig S, et al. PRIMAL: pathfinding via reinforcement and imitation multi-agent learning. *IEEE Robot Autom Lett.* 2019;4(3):2378–85. doi:10.1109/lra.2019.2903261.
 97. Chen Y, Rosolia U, Ames AD. Decentralized task and path planning for multi-robot systems. *IEEE Robot Autom Lett.* 2021;6(3):4337–44. doi:10.1109/lra.2021.3068103.
 98. Goarin M, Loianno G. Graph neural network for decentralized multi-robot goal assignment. *IEEE Robot Autom Lett.* 2024;9(5):4051–8. doi:10.1109/LRA.2024.3371254.
 99. Choi Y, Di Marco P, Park P. Communication-aware graph neural network for multi-agent reinforcement learning. *IEEE Access.* 2025;13:55832–40. doi:10.1109/access.2025.3554736.
 100. Cai Y, He X, Guo H, Yau WY, Lv C. Transformer-based multi-agent reinforcement learning for generalization of heterogeneous multi-robot cooperation. In: *Proceedings of the 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; 2024 Oct 14–8; Abu Dhabi, United Arab Emirates. p. 13695–702. doi:10.1109/IROS58592.2024.10802580.
 101. Forsberg AL, Nikou A, Feljan AV, Tumova J. Multi-agent transformer-accelerated RL for satisfaction of STL specifications. *arXiv:2403.15916*. 2024.

102. Liu Y, Chen J, Liu J, Jing X. Nonlinear mechanics of flexible cables in space robotic arms subject to complex physical environment. *Nonlinear Dyn.* 2018;94(1):649–67. doi:10.1007/s11071-018-4383-y.
103. Xia L, Sun Y, Liang C, Zhang L, Ali Imran M, Niyato D. Generative AI for semantic communication: architecture, challenges, and outlook. *IEEE Wirel Commun.* 2025;32(1):132–40. doi:10.1109/mwc.003.2300351.
104. Huang W, Abbeel P, Pathak D, Mordatch I. Language models as zero-shot planners: extracting actionable knowledge for embodied agents. *Comput Res Repos.* 2022;162:9118–47.
105. O'Neill A, Rehman A, Maddukuri A, Gupta A, Padalkar A, Lee A, et al. Open X-Embodiment: robotic learning datasets and RT-X models: open X-Embodiment collaboration⁰. *IEEE Int Conf Robot Autom.* 2024;2024(2):6892–903. doi:10.1109/ICRA57147.2024.10611477.
106. Kawaharazuka K, Oh J, Yamada J, Posner I, Zhu Y. Vision-language-action models for robotics: a review towards real-world applications. *IEEE Access.* 2025;13(140):162467–504. doi:10.1109/ACCESS.2025.3609980.
107. Zhao TZ, Kumar V, Levine S, Finn C. Learning fine-grained bimanual manipulation with low-cost hardware. *arXiv:2304.13705.* 2023.