SCIENCE CHINA
Information Sciences

• REVIEW •

May 2026, Vol. 69, Iss. 5, 151202:1–151202:31
https://doi.org/10.1007/s11432-025-4820-4

Multi-agent embodied AI: advances and future directions

Zhaohan FENG[1], Ruiqi XUE[2], Lei YUAN[2], Yang YU[2], Ning DING[3], Meiqin LIU[3,4], Bingzhao GAO[5], Jian SUN[1], Xinhu ZHENG[6] & Gang WANG[1*]

[1]*National Key Lab of Autonomous Intelligent Unmanned Systems, Beijing Institute of Technology, Beijing 100081, China*
[2]*National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China*
[3]*National Key Laboratory of Human-Machine Hybrid Augmented Intelligence, Xi'an Jiaotong University, Xi'an 710049, China*
[4]*College of Electrical Engineering, Zhejiang University, Hangzhou 310007, China*
[5]*Clean Energy Automotive Engineering Center, Tongji University, Shanghai 201804, China*
[6]*Intelligent Transportation Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511453, China*

Received 25 May 2025/Revised 19 September 2025/Accepted 2 November 2025/Published online 31 March 2026

Abstract Embodied artificial intelligence (embodied AI) plays a pivotal role in the application of advanced technologies in the intelligent era, where AI systems are integrated with physical bodies that enable them to perceive, reason, and interact with their environments. Through the use of sensors for input and actuators for action, these systems can learn and adapt based on real-world feedback, allowing them to perform tasks effectively in dynamic and unpredictable environments. As techniques such as deep learning (DL), reinforcement learning (RL), and large language models (LLMs) have matured, embodied AI has become a leading field in both academia and industry, with applications spanning robotics, healthcare, transportation, and manufacturing. However, most research has focused on single-agent systems that often assume static, closed environments, whereas real-world embodied AI must navigate far more complex scenarios. In such settings, agents must not only interact with their surroundings but also collaborate with other agents, necessitating sophisticated mechanisms for adaptation, real-time learning, and collaborative problem-solving. Despite increasing interest in multi-agent systems, existing research remains narrow in scope, often relying on simplified models that fail to capture the full complexity of dynamic, open environments for multi-agent embodied AI. Moreover, no comprehensive survey has systematically reviewed the advancements in this area. As embodied AI rapidly evolves, it is crucial to deepen our understanding of multi-agent embodied AI to address the challenges presented by real-world applications. To fill this gap and foster further development, this study reviews the current state of research, analyzes key contributions, and identifies challenges and future directions, providing insights to guide innovation and progress in this field.

Keywords embodied AI, multi-agent systems, reinforcement learning, cooperation and competition, generative model

Citation Feng Z H, Xue R Q, Yuan L, et al. Multi-agent embodied AI: advances and future directions. Sci China Inf Sci, 2026, 69(5): 151202, https://doi.org/10.1007/s11432-025-4820-4

1 Introduction

Embodied artificial intelligence (embodied AI) [1, 2] is an interdisciplinary research area at the intersection of artificial intelligence (AI), robotics, and cognitive science, aimed at equipping robots with capabilities for perception, planning, decision-making, and action, thereby enabling them to interact with and adapt to their environments actively. This concept was first introduced by Alan Turing in the 1950s, who explored how machines could perceive the world and make corresponding decisions [3]. Later, in the 1980s, researchers including Rodney Brooks reconsidered symbolic AI, suggesting that intelligence should emerge through active interaction with the environment rather than passive data-driven learning, thus laying the foundation for embodied AI [4]. In recent years, driven by advances in deep learning (DL), reinforcement learning (RL), and other techniques, embodied AI has made significant progress, particularly through the application of pre-trained models such as large language models (LLMs) and vision-language models (VLMs) [5]. These models have enhanced intelligent systems' capabilities in perception, decision-making, and task planning, achieving outstanding performance in visual understanding, natural language processing, and multimodal integration tasks [6]. Leveraging these large models, embodied AI has achieved breakthroughs in image recognition, speech interaction, and robotic control, thereby improving robots' ability to understand and adapt to complex dynamic environments [7–9].

* Corresponding author (email: gangwang@bit.edu.cn)

© Science China Press 2026

The primary challenge in embodied AI is that intelligent agents must simultaneously possess robust perception and decision-making capabilities, along with the capacity to continuously learn and adapt through ongoing interaction with dynamic and evolving environments [10]. Historically, early symbolic approaches, exemplified by Turing machine theory, attempted to achieve intelligence via symbolic representation and logical reasoning. However, these symbolic methods fell short in effectively addressing the dynamic interactions required between perception and action [11]. To overcome these limitations, Rodney Brooks proposed the concept of the "perception-action loop", which posits that intelligence emerges naturally through an agent's active and continuous engagement with its environment, thus laying the foundation for modern embodied AI research [12]. Following this line of thought, learning paradigms such as imitation learning (IL) emerged, accelerating the learning process by mimicking expert demonstrations, yet remaining heavily reliant on expert-generated data and consequently vulnerable to challenges like distributional shifts [13]. As an alternative, RL, especially deep RL, integrates deep neural networks to optimize decision-making strategies through interaction-driven feedback from the environment [14]. Such methods excel at managing high-dimensional sensory inputs, demonstrating strong adaptability across various complex tasks. More recently, the advent of large-scale models, particularly LLMs [15] and VLMs [16], has significantly advanced embodied AI, enhancing the capabilities of agents in visual understanding, multimodal perception, and task planning. Nevertheless, despite these advancements, most current research primarily addresses single-agent scenarios and often overlooks the inherent complexity of multi-agent interactions. In contrast, real-world embodied AI applications frequently demand effective collaboration among multiple heterogeneous agents to achieve tasks such as confronting unknown adversaries or addressing complex dynamic challenges [17, 18].

Multi-agent settings fundamentally differ from single-agent scenarios, as agents must simultaneously optimize their individual policies while coordinating with others. These differences give rise to several unique challenges.

(1) **Exponential action space.** The combined action space across multiple agents grows exponentially, significantly increasing the complexity of planning and decision-making.

(2) **Extended task horizons.** Inter-agent dependencies often require reasoning over longer time horizons to achieve effective coordination.

(3) **Partial observability.** Due to decentralized information, each agent has access to only a limited and potentially noisy local observation, complicating state estimation and policy learning.

(4) **Non-stationarity.** As multiple agents learn and adapt concurrently, the environment becomes inherently non-stationary from the perspective of any single agent, violating the standard Markov assumptions.

(5) **Credit assignment.** Determining the individual contribution of each agent to the overall team performance is often ambiguous, making it difficult to assign appropriate feedback for learning.

These challenges collectively differentiate multi-agent intelligence from its single-agent counterpart and necessitate the development of specialized algorithms and frameworks [19–22]. Despite substantial advancements in single-agent embodied AI, research on embodied AI within multi-agent contexts remains relatively nascent. Current research typically adapts successful single-agent methods or employs established frameworks such as RL and LLMs. Recently, the development of specialized benchmarks tailored explicitly to embodied multi-agent scenarios has begun, aiming to support systematic progress in this evolving field. While extensive literature reviews have thoroughly explored related domains, including embodied AI [1, 7], multi-agent reinforcement learning (MARL) [19, 21], and multi-agent cooperation [20, 23], comprehensive surveys explicitly focusing on embodied multi-agent AI remain limited. For example, Ref. [24] synthesized embodied MAS frameworks, mechanisms, and approaches by drawing an analogy to control system design methodologies, thereby providing a systematic classification scheme. Complementing this control-oriented view, Ref. [25] offered a learning-centered perspective, systematically reviewing recent progress in embodied MARL, including social learning, emergent communication, Sim2Real transfer, hierarchical methods, and safety considerations. Ref. [5] further extended this perspective by reviewing the integration of generative foundation models into embodied multi-agent systems (MAS), proposing a taxonomy of collaborative architectures, and discussing essential components like perception, planning, communication, and feedback mechanisms. Nevertheless, these surveys primarily address specific dimensions of multi-agent embodied AI, lacking a systematic and comprehensive overview of the entire field.

Recognizing the substantial potential of multi-agent embodied AI for addressing complex coordination tasks in real-world environments, this study provides a systematic and comprehensive review of recent advances in this emerging research area. As shown in Figure 1, we first introduce foundational concepts, including MAS, RL, and relevant methodologies. Next, we discuss embodied AI within single-agent contexts, clearly outlining core definitions, primary research directions, representative methods, and established evaluation benchmarks. Building on this foundation, we extend our discussion to multi-agent embodied AI, highlighting widely employed techniques and examining recent prominent benchmarks specifically designed for multi-agent scenarios. Finally, we present insightful perspectives on the future development of multi-agent embodied AI and aim to stimulate further research

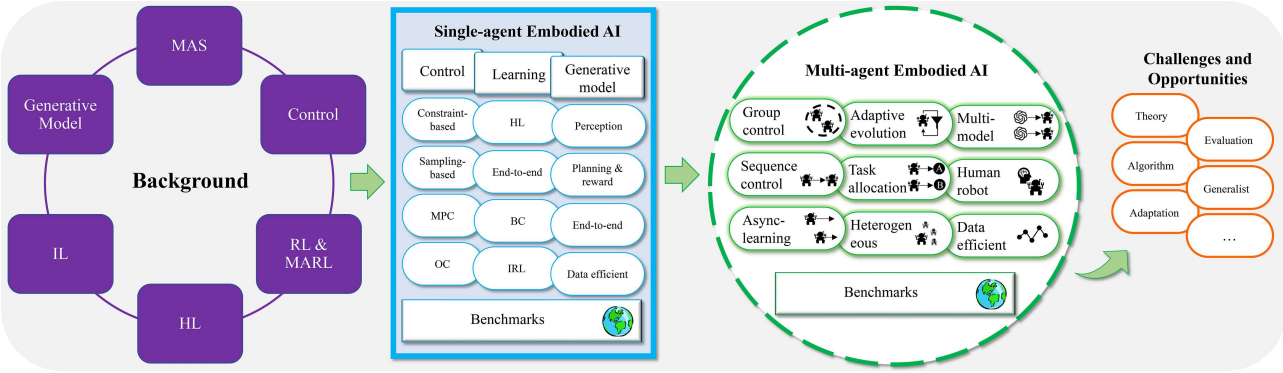


Figure 1 (Color online) An overview of this survey.

and innovation in this promising and rapidly evolving field. To facilitate ease of reading and reference, the key contributions of this survey are summarized as follows.

• **Systematic background knowledge.** We provide a comprehensive overview of the foundations of embodied AI and multi-agent systems in Section 2, covering core definitions, conceptual frameworks, and representative modeling paradigms. This section offers a structured review of the theoretical and algorithmic underpinnings essential for understanding multi-agent embodied AI.

• **Taxonomy of core methods.** We propose a unified taxonomy that categorizes mainstream approaches into three technical routes: control and planning, learning-based methods, and generative model-driven frameworks. These routes are consistently analyzed under both single-agent (Section 3) and multi-agent (Section 4) embodied settings.

• **Benchmark landscape summary.** We present a detailed overview of widely adopted benchmarks across both single-agent and multi-agent embodied AI tasks. These benchmarks span navigation, manipulation, human-AI collaboration, multimodal interaction, and simulation-to-real transfer. Subsections 3.4 and 4.4 provide structured benchmark tables about both single-agent settings and multi-agent settings to facilitate rapid lookup and comparison.

• **Future research directions.** We identify and discuss key open challenges and promising research directions in multi-agent embodied AI, including theoretical modeling, algorithmic scalability, data-efficient learning, and foundation model integration. Section 5 outlines 8 thematic areas that are critical for future progress. This guidance aims to help researchers pinpoint impactful directions in this rapidly evolving field.

2 Preliminaries

In this section, we present the core techniques that underpin embodied AI, starting with a formal definition of embodied AI itself. We also define the concept of MAS and explore a range of essential methods, including optimal control (OC), RL, MARL, hierarchical learning, IL, and generative models.

2.1 Embodied AI

Embodied AI (Figure 2) refers to a class of intelligent agents equipped with physical bodies that enable them to perceive, act upon, and adapt to the environments through continuous interaction [7]. The concept of embodied AI traces back to Alan Turing's early proposition in the 1950s, suggesting that genuine intelligence must arise from sensory and motor experiences rather than purely symbolic computation [3]. This notion was further formalized in the 1980s through embodied cognition theory, which posits that cognition is inherently shaped by the agent's physical form and interaction with the world [4]. In contrast to traditional AI paradigms that rely on abstract reasoning or passive learning from static datasets, embodied AI emphasizes real-world interaction as the foundation for learning and decision-making.

At a system level, embodied AI architectures are typically composed of three tightly integrated components: perception, cognition, and action. Firstly, agents use sensors to acquire real-time, context-dependent information from their surroundings. Next, the sensory data is processed by cognitive modules that support reasoning, interpretation and planning. Finally, resulting decisions are translated into actions via actuators, which modify the environment and initiate new perceptual inputs. These processes form a continuous feedback loop known as the perception-cognition-action cycle [26], which enables embodied agents to adapt their behavior dynamically based on

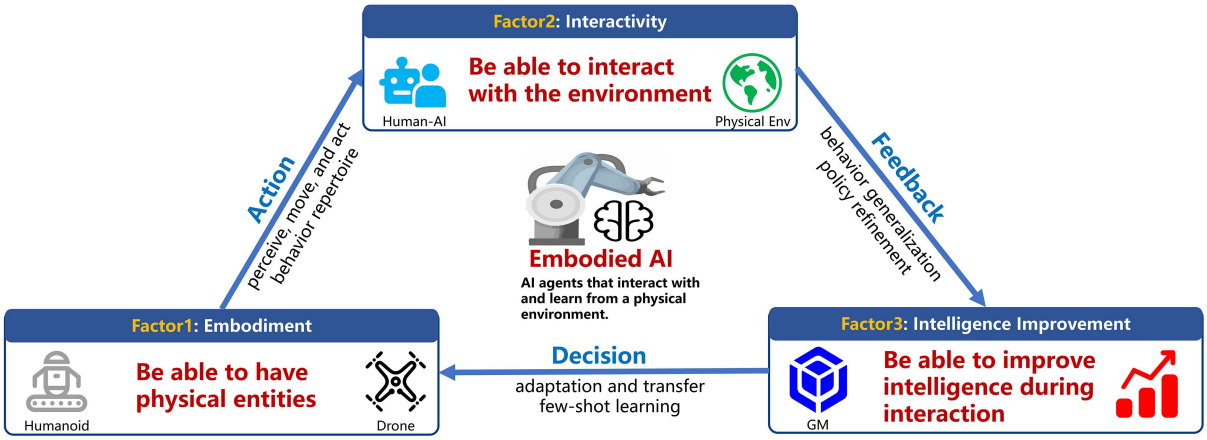


Figure 2 (Color online) An illustration of embodied AI.

feedback. The paradigm of embodied AI mainly includes three foundational attributes that govern how intelligence emerges and evolves.

• **Embodiment.** Embodied AI is rooted in physical agents equipped with the capabilities to perceive, move, and act in the real world. These agents come in diverse forms, including humanoid robots, quadrupeds, unmanned vehicles and drones. The physical body serves not only as the medium through which the agent interacts with its environment, but also as the structural foundation that constrains and enables its behavioral repertoire. The morphology, sensorimotor fidelity and actuation of the body jointly define the range and granularity of the agent's interactions, shaping the scope of its situated intelligence.

• **Interactivity.** Built upon this physical grounding, embodied intelligence emerges through continuous, closed-loop interaction with the surrounding environment. Operating from a first-person perspective, the agent engages in a dynamic cycle of perception, decision-making, and action. Each behavior not only responds to environmental stimuli but also alters future sensory inputs, forming a feedback-rich loop that supports adaptive learning. Through this persistent engagement, the agent refines its policy, acquires task-specific competencies, and generalizes behaviors across varying contexts, allowing for robust, context-aware performance in real-world scenarios.

• **Intelligence improvement.** The development of embodied AI is characterized by a capacity for continual improvement in both cognition and behavior. This progression is increasingly enabled by the integration of large-scale multimodal models, which endow agents with semantic understanding, instruction following, and contextual reasoning. These models facilitate few-shot learning, in-context adaptation, and knowledge transfer across tasks. As the agent interacts with its environment, it incrementally aligns its perceptual inputs, decision processes, and physical actions, leading to both immediate task success and sustained growth in autonomy, adaptability, and generalization over time.

Recent advancements in generative models such as LLMs [27] have further expanded the cognitive capabilities of embodied agents. Through leveraging their powerful reasoning and generalization abilities, LLMs enable embodied agents to understand language instructions, ground semantic knowledge in physical experiences, and perform zero-shot or few-shot adaptations. These developments have accelerated the deployment of embodied AI in various real-world domains such as robotics, autonomous driving, intelligent manufacturing and healthcare [16]. Importantly, embodied AI is not merely the integration of powerful AI models with robotic platforms. Instead, it represents a co-evolutionary paradigm where intelligent algorithms which serve as brains, physical structures which serve as bodies and dynamic environments evolve jointly to support adaptive embodied intelligence.

2.2 Multi-agent systems

MAS are composed of multiple autonomous agents, each capable of perceiving its environment, making independent decisions and executing actions accordingly [17]. In contrast to centralized control paradigms, MAS adopt a decentralized architecture in which agents interact locally while achieving global coordination. This decentralized design offers notable advantages in scalability, fault tolerance, and adaptability, particularly in dynamic, partially observable, or non-stationary environments. Depending on the nature of agent interactions, MAS can be broadly categorized into cooperative, competitive, and mixed settings, as illustrated in Figure 3. Core attributes of MAS include autonomy, decentralization, inter-agent communication, local information access, and dynamic adaptability. These features collectively empower MAS to tackle a broad spectrum of complex, high-dimensional tasks that

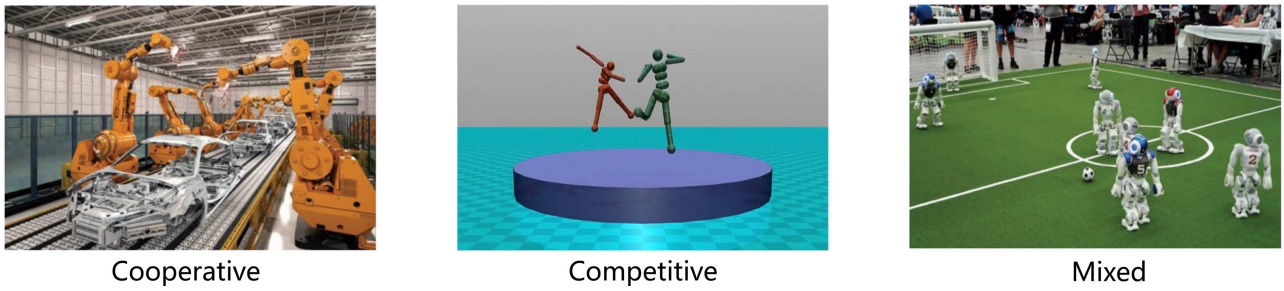


Figure 3 (Color online) Three common settings of MAS.

demand parallel sensing, distributed planning, and real-time coordination, with prominent applications in domains such as robotics, autonomous driving, and smart infrastructure.

In recent years, MAS research has undergone a significant paradigm shift, fueled by the convergence of learning-based methods and advances in neural architectures. At the forefront of this transformation is MARL [21], which provides a powerful framework for enabling agents to learn complex behaviors through interaction. Techniques such as centralized training with decentralized execution (CTDE), parameter sharing, credit assignment, and opponent modeling have been widely adopted to address core challenges including non-stationarity, coordination, and partial observability. Complementing these advances, the integration of LLMs has opened up new capabilities for MAS. LLM-enabled agents can access vast pretrained knowledge, communicate via natural language, and engage in high-level reasoning and abstraction, capabilities that transcend the limitations of conventional policy-driven systems. The fusion of reinforcement learning and foundation models is thus reshaping the MAS landscape, paving the way for more generalizable, interpretable, and human-aligned agent architectures [28].

2.3 Optimal control

OC is a fundamental branch of modern control theory that focuses on designing control policies to optimize a specific performance criterion, such as time, energy consumption, or cost, while satisfying dynamic and control constraints of the system [29]. Owing to its ability to achieve global optimality, handle multiple objectives and constraints simultaneously, and maintain robustness in complex systems, OC has found wide-ranging applications in fields such as aerospace engineering, industrial automation, economic management, and energy systems.

Formally, a standard OC problem can be described as follows. Given a dynamic system governed by $\dot{x}(t) = f(x(t), u(t), t)$ and an initial state $x(t_0) = x_0$, the goal is to find the optimal control input $u(t)$ that minimizes the performance index:

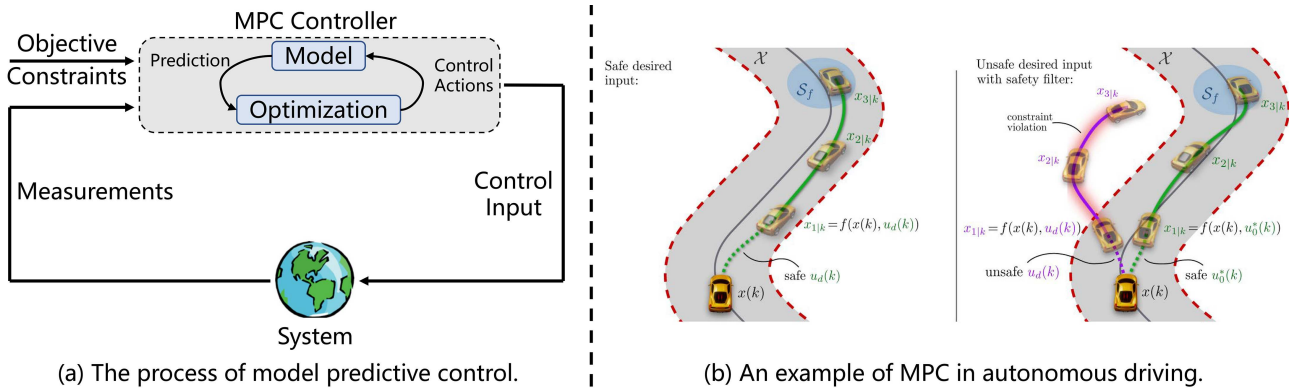
$$J = \phi(x(t_f), t_f) + \int_{t_0}^{t_f} L(x(t), u(t), t)dt, \tag{1}$$

where ϕ is the terminal cost function and L is the running cost function.

Common methods in OC include the calculus of variations, Pontryagin's maximum principle, dynamic programming, and model predictive control (MPC). The calculus of variations derives optimal trajectories using the Euler-Lagrange equations. PMP formulates necessary conditions for optimality by constructing a Hamiltonian function and defining both state and costate dynamics. DP addresses state-space optimization in complex systems via Bellman's principle of optimality. In contrast, MPC solves constrained optimization problems in real time using a receding horizon strategy. Among them, MPC is the most widely adopted in industry. Its core idea is to solve a finite-horizon OC problem at each time step and implement only the first OC action, as shown in Figure 4. Specifically, at time t_k, MPC solves the following online optimization problem:

$$\min_{u_0, u_1, \ldots, u_{N-1}} \sum_{i=0}^{N-1} L(x_i, u_i) + \phi(x_N), \tag{2}$$

subject to the system dynamics $x_{i+1} = f(x_i, u_i, i)$. Due to its efficiency, real-time capability, and robustness, MPC has been widely applied in embodied AI systems with multiple agents, such as coordinated control of robot swarms and formation control of unmanned aerial vehicles (UAVs).



(a) The process of model predictive control. (b) An example of MPC in autonomous driving.

Figure 4 (Color online) An illustration of MPC.

2.4 Reinforcement learning

Reinforcement learning (RL) is a fundamental paradigm of machine learning concerned with enabling agents to make sequential decisions by interacting with dynamic environments [30]. Unlike supervised learning, which requires labeled data, or unsupervised learning, which uncovers hidden patterns in unlabeled data, RL is grounded in a trial-and-error process. Through continuous interaction, agents learn to select actions that maximize long-term cumulative rewards, using feedback from the environment to improve their behavior over time. Owing to its capacity for learning from interaction, RL has achieved impressive results across a broad range of domains. These include strategic game playing [31,32], robotic control [33], industrial process optimization [34], personalized healthcare [35], autonomous driving [36,37], and even alignment and instruction-following in LLMs [38,39].

At its core, an RL problem is commonly formalized as a Markov decision process (MDP), depicted in Figure 5(a). An MDP is defined as a tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, P, R, \gamma \rangle$, where $\mathcal{S}$ and $\mathcal{A}$ represent the sets of states and actions, $P(s'|s,a)$ is the transition probability function, $R(s,a)$ denotes the reward function, and $\gamma \in [0,1]$ is the discount factor that balances short- and long-term rewards. The objective is to learn a policy $\pi(a|s)$ that maximizes the following expected cumulative return

$$\mathbb{E}_\pi\Big[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \mid s_0\Big], \tag{3}$$

where the expectation $\mathbb{E}_\pi[\cdot]$ is taken over trajectories generated by the policy π and the transition dynamics P. While single-agent RL provides a solid foundation for autonomous decision-making, many real-world scenarios involve multiple agents acting and learning simultaneously. This motivates the extension of RL into MARL, where coordination or competition among agents becomes essential. Specifically, in fully cooperative tasks, MARL problems are often modeled as the decentralized partially observable Markov decision process (Dec-POMDP) framework as shown in Figure 5(b). A Dec-POMDP extends the MDP by introducing multiple agents and partial observability. Formally, it is defined as a tuple $\mathcal{M} = \langle I, \mathcal{S}, \mathcal{A}_i, P, R, \Omega_i, O, n, \gamma \rangle$, where I is the set of n agents. Each agent i selects actions from its own action space $\mathcal{A}_i$, based solely on local observations from Ω_i. The global state transitions according to the joint action $\boldsymbol{a}$ via $P(s'|s,\boldsymbol{a})$, and all agents receive a shared reward $R(s,\boldsymbol{a})$. Observations are generated through the function $O(s',\boldsymbol{a},\boldsymbol{o})$, providing each agent with private, potentially incomplete information. The central challenge lies in learning a set of decentralized policies $[\pi_1, \ldots, \pi_n]$, where each agent must act appropriately using local information, yet their joint behavior maximizes the team's cumulative reward.

A wide range of algorithmic strategies has been developed to address both single-agent and multi-agent RL problems [40–42]. Among these, two principal categories stand out: value-based methods and policy-based methods. Value-based methods, such as the deep Q-network (DQN) [43], focus on estimating action-value functions, enabling agents to act greedily with respect to predicted future rewards. In contrast, policy-based methods, such as DDPG [44] and PPO [45], directly optimize parameterized policies. These methods often incorporate value functions to guide learning and are especially suited to environments with high-dimensional or continuous action spaces. Both paradigms have been extended to the multi-agent settings through distinct architectural innovations. Value decomposition methods, such as QMIX [46], learn a global value function that can be decomposed into individual agent utilities. This structure supports decentralized execution while enabling centralized training. On the other hand, policy-based MARL approaches, including MADDPG [47] and MAPPO [48], adopt the CTDE framework. These methods employ a shared global critic during training to provide stabilized learning signals and support effective credit assignment among agents, thereby enhancing coordination in cooperative tasks.

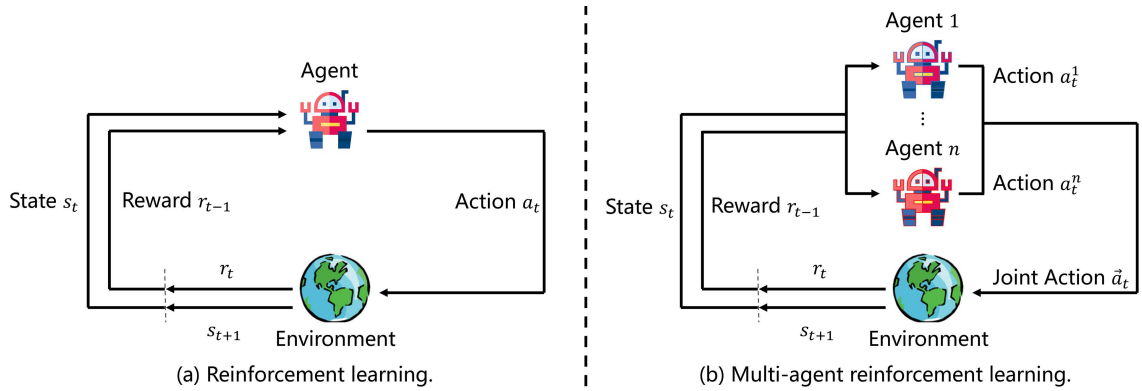


Figure 5 (Color online) An illustration of RL and MARL.

2.5 Hierarchical learning

Hierarchical learning is a paradigm that organizes the learning process into multiple levels of abstraction, with higher levels responsible for setting abstract goals or intentions, and lower levels focusing on executing more concrete, fine-grained subtasks [49]. This layered structure enables models to operate across varying levels of granularity, thereby improving efficiency and scalability in solving complex tasks by breaking them down into simpler, manageable components [50].

The core process of hierarchical learning typically involves two stages: low-level skill learning and high-level policy learning. The low-level skill learning is aimed at mastering basic sub-tasks and is often implemented using classical controllers such as MPC [51–53], or through end-to-end RL [54, 55]. In contrast, the high-level policy learning is responsible for orchestrating the learned low-level skills to accomplish more complex goals. This is commonly achieved through RL or via planning with LLMs. In RL-based high-level policy learning, the action space is defined as the set of learned low-level policies. A parameterized high-level policy is then trained using reward signals from the environment to select and sequence these low-level skills effectively [54, 55]. In contrast, LLM-based high-level policy learning typically involves providing the task objective and the available set of low-level skills as input to the LLM, which then directly generates a structured plan by composing and invoking the appropriate low-level skills to accomplish the task [49].

2.6 Imitation learning

IL, also referred to as learning from demonstrations, is a machine learning paradigm in which an agent acquires task-solving capabilities by observing and mimicking expert behavior. In contrast to RL, which requires the manual design of a reward function to guide learning, IL leverages expert demonstrations as direct supervision. This distinction makes IL particularly advantageous in complex, high-dimensional environments where reward specification is ambiguous, costly, or even infeasible. Over the years, three approaches have emerged as the most prominent within IL: behavior cloning (BC), inverse reinforcement learning (IRL), and generative adversarial imitation learning (GAIL) [13].

Among these, BC is the most straightforward and widely adopted approach. It casts the imitation task as a supervised learning problem, aiming to directly map observed states to corresponding expert actions. As illustrated in Figure 6(a), the agent learns from a dataset of expert-generated state-action pairs $\mathcal{D} = \{(s_i, a_i)\}_{i=1}^N$, optimizing the following objective:

$$\min_\theta \mathbb{E}_{(s,a)\sim\mathcal{D}}[-\log \pi_\theta(a|s)], \tag{4}$$

where θ denotes the parameters of the policy π_θ, this loss encourages the learned policy to produce actions that closely match those of the expert when exposed to similar states. While BC is computationally efficient and simple to implement, it suffers from compounding error due to covariate shift, especially in long-horizon tasks where the agent visits states not seen during training.

To address some of BC's limitations, IRL adopts a fundamentally different approach by aiming to infer the expert's underlying reward function rather than replicating their actions directly. Once the reward function is recovered, standard RL methods can be used to derive the optimal policy. This two-step process, illustrated in Figure 6(b), enables improved generalization and policy interpretability. Formally, a representative formulation is

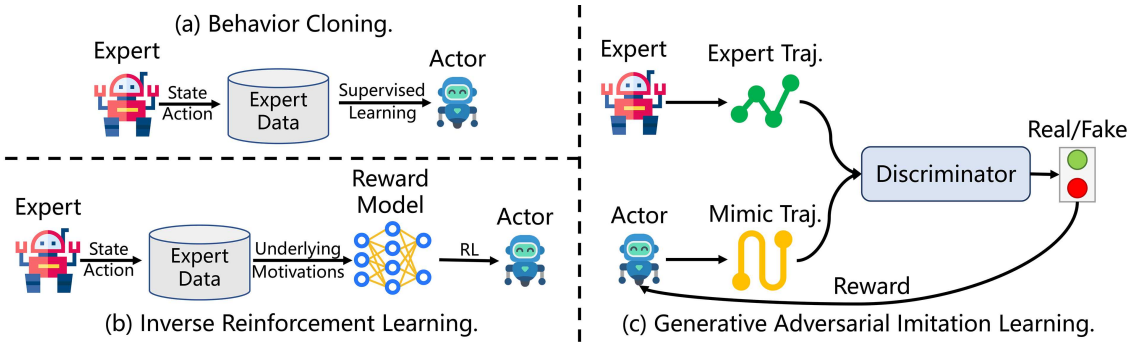


Figure 6 (Color online) An illustration of different IL methods.

the maximum entropy IRL objective

$$\max_{R} \mathbb{E}_{\tau\sim\mathcal{D}}[R(\tau)] - \log Z(R), \tag{5}$$

where $\tau = (s_0, a_0, \ldots, s_T, a_T)$, $R(\tau) = \sum_{t=0}^{T} R(s_t, a_t)$, and $Z(R) = \int e^{R(\tau)} d\tau$ is the partition function that ensures a normalized trajectory distribution. By explicitly modeling the expert's intent, IRL provides more robust and explainable policies. However, this benefit comes at the cost of additional computational overhead due to the requirement for repeated RL iterations during reward learning.

Different from both BC and IRL, GAIL [56] introduces an adversarial training framework that eliminates the need for explicit reward specification. Inspired by generative adversarial networks (GANs) [57], GAIL frames imitation as a minimax game between two components: a generator (the policy to be learned) and a discriminator that distinguishes expert trajectories from those generated by the policy. As depicted in Figure 6(c), the optimization objective is

$$\min_{\pi} \max_{D} \mathbb{E}_{(s,a)\sim\mathcal{D}}[\log D(s,a)] + \mathbb{E}_{(s,a)\sim\mathcal{D}_\pi}[\log(1 - D(s,a))], \tag{6}$$

where $\mathcal{D}_\pi$ represents trajectories generated by the current policy. Through this adversarial setup, GAIL implicitly learns a reward signal that encourages the agent to behave indistinguishably from the expert. Compared to IRL, GAIL streamlines the training process by avoiding explicit reward modeling, resulting in better sample efficiency. However, this advantage comes at the expense of reduced interpretability, as the learned reward is never explicitly extracted. In summary, BC, IRL, and GAIL represent complementary paradigms within the broader IL framework. BC provides a simple and effective baseline, IRL offers interpretability and generalization via reward recovery, and GAIL combines the strengths of both by enabling end-to-end imitation through adversarial learning. The choice among them often depends on the trade-offs between sample efficiency, generalization capability, computational complexity, and interpretability required by the target application.

2.7 Generative models

Generative models constitute a foundational paradigm in machine learning, with the central objective of capturing the underlying distribution of training data to generate novel samples that exhibit similar characteristics. These models have found widespread applications across domains such as vision, language and multimodal learning. In recent years, the emergence of large-scale generative models, such as LLMs and VLMs, has significantly advanced the field. Their success is largely attributed to strong generalization capabilities, the availability of massive datasets and scalable architectures. At the core of these models lie several key architectural frameworks, including Transformers [58], diffusion models [59] and more recently, state space models (SSMs) such as Mamba [60]. Among these, the Transformer architecture has played a pivotal role in revolutionizing sequence modeling. Originally proposed for machine translation, Transformers eliminate the need for recurrence or convolution by introducing an attention-based mechanism that enables each element in a sequence to directly attend to every other element (cf. Figure 7(a)). This design facilitates efficient parallel computation and allows the model to capture global contextual dependencies. The core attention mechanism is mathematically defined as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right) V, \tag{7}$$

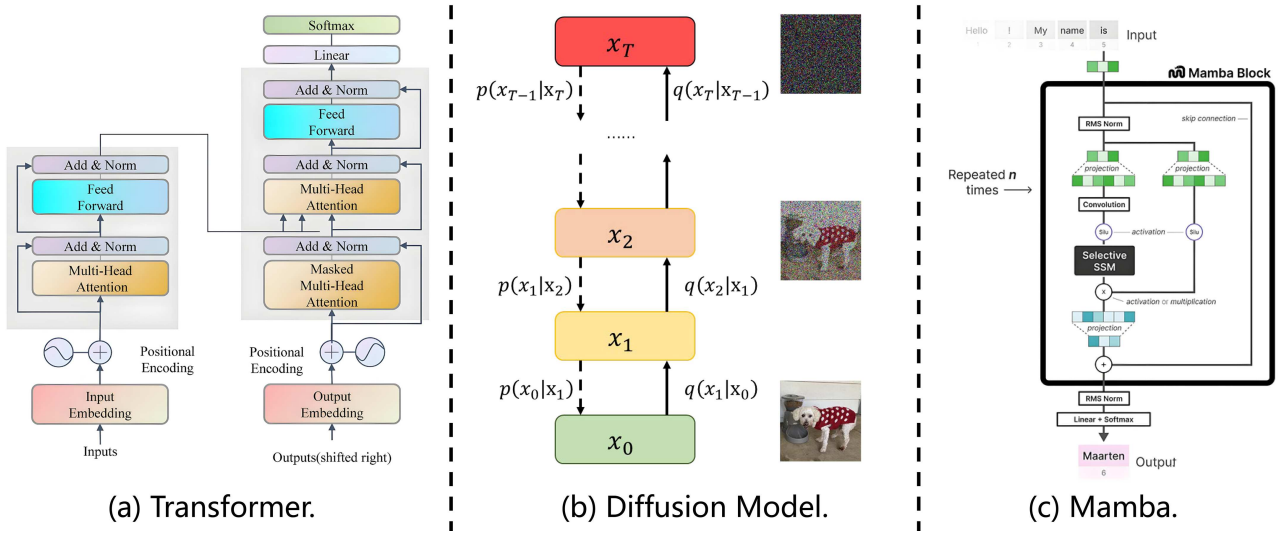


Figure 7 (Color online) An illustration of different generative model structures.

where Q, K, and V represent the queries, keys, and values, respectively, and d_k is the dimensionality of the key vectors, acting as a scaling factor to stabilize gradients. In the case of self-attention, Q, K, and V are computed from the same input sequence, enabling each token to contextualize itself with respect to all others. This mechanism is fundamental to the success of Transformers in a wide range of natural language processing and vision tasks.

In contrast to the discrete attention mechanisms of Transformers, diffusion models offer a probabilistic, noise-driven framework for generation. These models operate through a two-phase process: a forward diffusion stage that incrementally corrupts data with noise, and a learned reverse process that reconstructs the data from noisy inputs (cf. Figure 7(b)). The forward diffusion process forms a Markov chain in which Gaussian noise is added at each timestep:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I), \tag{8}$$

where β_t denotes a predefined variance schedule and I is the identity matrix. The reverse process, which the model learns to approximate, also follows a Gaussian distribution:

$$p(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)), \tag{9}$$

where μ_θ and Σ_θ are parameterized by neural networks, typically based on U-Net architectures [61]. By iteratively denoising from pure Gaussian noise, diffusion models can generate high-fidelity outputs, making them highly effective for image generation, audio synthesis, and other structured data domains.

While Transformers and diffusion models have achieved remarkable success, they both encounter limitations when scaling to long sequences or complex generative processes. To address these challenges, the Mamba architecture introduces a novel approach based on continuous-time SSMs, offering linear-time complexity and improved efficiency in long-sequence processing (cf. Figure 7(c)). Mamba represents sequential data using the following differential equations:

$$\dot{h}(t) = A(t)h(t) + B(t)x(t), \;\; y(t) = C(t)h(t) + D(t)x(t), \tag{10}$$

where $h(t)$ is the hidden state, $x(t)$ and $y(t)$ denote the input and output signals, respectively, and $A(t)$, $B(t)$, $C(t)$, and $D(t)$ are learnable, time-dependent system parameters. In most practical scenarios, $D(t)$ is set to zero, introducing a skip connection that improves gradient flow. By discretizing the continuous system and introducing a selective scanning mechanism, Mamba achieves superior inference efficiency and scalability. These properties make it a compelling alternative for tasks involving long-range dependencies and real-time computation. Taken together, Transformers, diffusion models, and SSMs like Mamba represent three major architectural directions in generative modeling. Each of these paradigms provides unique advantages in terms of expressiveness, efficiency, and applicability, and understanding their design principles is critical for the development of next-generation generative systems.

Table 1 Methods for building embodied AI in single-agent settings, where the abbreviation GM refers to generative model.

	Year	Category			Input				View		Agent type					Implement		
		Perception	Planning	Control	Image	Point cloud	Language	Proprioception	Global	Local	Humanoid	Quadruped/Bipedal	UAV/UGV/USV	Manipulator	No certain type	Solver-based	NN-based	GM-based
RRT [62]	1998		✓					✓	✓						✓	✓		
MINCO [63]	2022		✓	✓				✓	✓				✓			✓		
High-MPC [51]	2022		✓	✓				✓	✓				✓			✓	✓	
SteveEye [64]	2023	✓	✓		✓		✓			✓					✓			✓
SayCan [49]	2023	✓	✓	✓	✓		✓			✓			✓	✓			✓	✓
VisionRacing [65]	2023	✓		✓	✓			✓		✓			✓				✓	
GNFactor [66]	2023	✓	✓	✓	✓		✓	✓		✓				✓			✓	
PaLM-E [67]	2023	✓	✓		✓		✓	✓		✓			✓	✓				✓
InnerMonologue [68]	2023	✓	✓		✓		✓			✓			✓	✓				✓
Infer & Adapt [69]	2024			✓				✓		✓		✓					✓	
RT-H [70]	2024	✓	✓	✓	✓		✓			✓			✓	✓				✓
OpenVLA [71]	2024	✓	✓	✓	✓		✓			✓			✓	✓				✓
Octo [72]	2024	✓	✓	✓	✓		✓	✓		✓					✓		✓	✓
$\pi_{0.5}$ [73]	2025	✓	✓	✓	✓		✓	✓		✓			✓	✓				✓

3 Single-agent embodied AI

The mainstream research in embodied AI predominantly focuses on the single-agent settings, where an individual agent operates in a static environment to complete various tasks. In this context, existing approaches can be broadly categorized into classical methods and modern learning-based techniques, as summarized in Table 1. Early efforts primarily relied on knowledge-driven, task-specific planning and control strategies [74–76]. However, as agents are increasingly deployed in dynamic and unstructured environments, manually designing bespoke planners and controllers for each scenario has become infeasible. Thus, learning-based methods with inherent adaptability and strong generalization across diverse tasks have been the prevailing paradigm for instilling embodied agents with intelligence in complex, dynamic environments [77–79]. More recently, advances in generative models have further enhanced the potential of embodied agents, enabling them to exhibit more adaptive and broadly applicable behaviors.

3.1 Classical control and planning

To accomplish long-horizon tasks in the physical world, embodied agents must effectively plan and control their motions, generating trajectories grounded in both perceptual and task-specific inputs. To this end, a rich body of classical planning methods has been developed, which can be broadly grouped into constraint-based, sampling-based, and optimization-based approaches [80]. Constraint-based methods encode task objectives and environmental conditions as logical constraints, translating the planning domain into symbolic representations, and employ constraint-solving techniques, such as symbolic search, to identify feasible solutions [81,82]. However, these methods typically focus on feasibility rather than solution quality, often neglecting optimality. Additionally, when perceptual inputs are high-dimensional or structurally complex, the difficulty of constraint solving increases substantially. Sampling-based methods address these limitations by incrementally exploring the feasible solution space through random sampling techniques, such as rapidly-exploring random trees (RRT) and its variants, progressively building a tree or graph structure to discover viable motion trajectories [62, 83–85].

To further optimize motion planning within the feasible domain, optimization-based methods explicitly model task objectives and performance metrics as objective functions, while representing feasibility conditions as hard constraints. Advanced optimization techniques are then utilized to search for optimal solutions within the constrained solution space. Representative optimization-based methods include polynomial trajectory planning [86], MPC [87] and OC [85]. Among these, polynomial trajectory planning generally exhibits relatively low computational cost, ensuring real-time responsiveness and smooth trajectory generation, making it particularly suitable for autonomous navigation tasks in complex scenarios [88–92]. In contrast, MPC and OC approaches are better

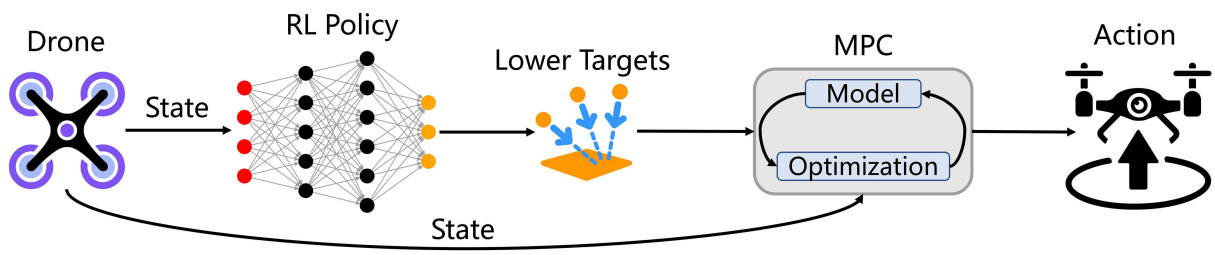


Figure 8 (Color online) An illustration of hierarchical learning using the example of embodied drone control.

suited for highly constrained scenarios requiring time-optimal maneuvers, such as collision avoidance during UAV operations [85, 93–95]. For example, the work [95] introduces an aggressive collision-inclusive motion planning framework that strategically utilizes collisions to improve navigation efficiency in constrained environments. This method integrates an enhanced collision-inclusive rapidly-exploring random tree (ICRRT) for generating deformable motion primitives, alongside a trajectory optimization approach guided by a receding horizon decision strategy for autonomous collision avoidance. Experimental results demonstrate a substantial reduction in task completion time compared to conventional collision-free methods, showing the effectiveness of this approach and underscoring its considerable potential for practical applications.

3.2 Learning-based methods

Classical control and planning methods have long been employed for constrained real-time decision-making due to their ability to produce high-precision solutions. However, these approaches are often computationally intensive, limiting scalability and responsiveness, especially in high-dimensional, nonlinear or non-stationary systems. Moreover, they typically exhibit poor generalization to unseen scenarios or dynamic environments. To overcome these limitations, learning-based decision-making paradigms have gained increasing attention, aiming to deliver real-time performance with improved robustness and generalization by learning directly from interactive data [96]. Among these, end-to-end RL has emerged as a prominent approach as it enables direct policy optimization through interaction with the environment. Nonetheless, in tasks with complex objectives or large state-action spaces, end-to-end RL suffers from low sample efficiency and prolonged training times due to the expansive exploration space. To alleviate this issue, hierarchical RL has been introduced, which decomposes overall tasks into a sequence of sub-tasks, thereby improving learning efficiency and scalability. Despite these advances, both end-to-end and hierarchical RL approaches fundamentally depend on the availability of well-designed reward functions, which are often difficult or impractical to define in real-world settings. In such scenarios, IL provides a compelling alternative by enabling agents to acquire effective policies directly from expert demonstrations, circumventing the need for explicit reward engineering and facilitating faster deployment in practice.

End-to-end learning. In sophisticated environments, OC problems often exhibit non-convex characteristics, making it challenging for classical planning and control methods to identify feasible solutions within a reasonable computational time. A natural approach to addressing this issue is to optimize the entire system in an end-to-end manner, employing RL techniques to directly map perceptual information to action decisions through neural networks. These methods optimize the policy by maximizing the expected cumulative rewards as defined in (3), ensuring stable and high-frequency control outputs [97]. Consequently, end-to-end learning is widely recognized as an ultimate pathway toward embodied AI, as it removes the necessity for manual decomposition, alignment, and tuning of individual modules, enabling agents to directly translate raw sensory inputs into effective actions. Indeed, end-to-end RL has achieved impressive performance on diverse robotic platforms equipped with multimodal inputs, including UAVs [65, 97], autonomous vehicles [98], manipulators [99] and legged robots [100, 101].

Hierarchical learning. When dealing with tasks characterized by complex objectives or large state-action spaces, end-to-end RL often faces challenges related to low sample efficiency and prolonged training times due to expansive exploration spaces. To address this issue, researchers frequently adopt hierarchical frameworks integrating RL with classical control methods. In such frameworks, RL typically manages high-level planning or policy generation, while robust low-level controllers such as MPC or proportional integral derivative (PID) controllers execute precise and reliable action commands [51, 52, 102]. This hierarchical approach enables embodied agents to effectively tackle challenging unstructured problems. For example, guiding a UAV through a dynamically swinging gate in minimal time is particularly difficult for classical methods alone (Figure 8), as the gate's position evolves over time, making it impractical to define a fixed optimization objective. Instead, RL can learn policies that determine intermediate target states and planning horizons for MPC, using reward signals shaped by traversal success and

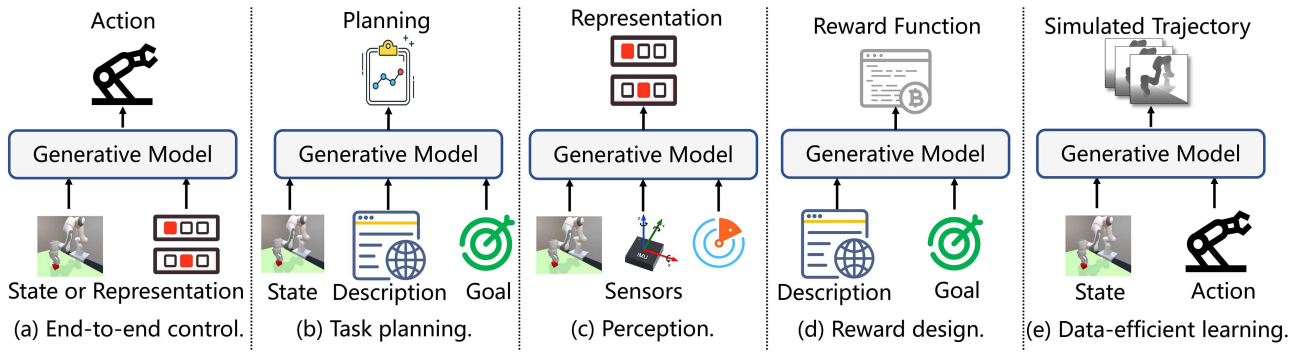


Figure 9 (Color online) An illustration of different roles generative models played in embodied control.

timing to generate efficient trajectories [51]. Furthermore, when gate dynamics are unknown, a learnable mixing coefficient can dynamically adjust the control policy, enabling smooth transitions between pursuit and passage behaviors based on real-time state information, thereby enhancing trajectory continuity [52]. Additionally, carefully structured hierarchical designs can naturally extend this framework to support multi-agent collaborative tasks in complex embodied scenarios [102].

Learning from demonstrations. Unlike OC methods, learning-based approaches such as RL require extensive data collection through interaction with the environment in a trial-and-error manner, incurring high costs in real-world embodied settings. Furthermore, RL often relies heavily on meticulously designed objective functions or reward signals, which become challenging to specify accurately in complex and unstructured environments due to the necessity for flexibility and adaptability. In these scenarios, learning from expert demonstrations, commonly referred to as IL, presents a more practical alternative. For example, Ref. [103] successfully transferred a painting policy from a simulator to a real-world robotic arm using BC, demonstrating high-level painting and calligraphy skills. Beyond this, a range of more advanced IL techniques have been introduced to further enhance embodied intelligence with improved generalization and robustness, including interactive IL [104], constrained IL [105], and IRL [106].

3.3 Generative model-based methods

Despite their progress, traditional learning methods often suffer from limited representational capacity, making them inadequate for handling complex scenarios such as multi-goal or multi-task decision making. These methods typically struggle with poor learning performance, inefficient use of external prior knowledge and low sample efficiency. To address these challenges, researchers have begun exploring the use of generative models, which offer stronger representational capabilities, to enable more efficient and flexible embodied intelligence [98,107]. The most straightforward application of generative models is to serve as end-to-end controllers by directly outputting executable actions, thereby transferring their internal prior knowledge into embodied agents. However, many embodied tasks involve fine-grained low-level behaviors which generative models often struggle to understand and control effectively. Inspired by the principles of hierarchical learning, a growing body of work has adopted generative models for generating high-level plans, while relying on traditional control or learning-based methods to execute low-level skills. Beyond direct decision-making, generative models are also widely used to support the learning process. Applications include integrating sensory information for perception, generating reward functions, and synthesizing data to improve sample efficiency.

End-to-end control. To enhance the utilization of prior knowledge and the processing of complex, multimodal observations during decision-making, a growing body of research has explored using generative models, particularly large VLMs, as direct decision-makers, as shown in Figure 9(a). Early approaches typically involve formatting inputs such as states and goals in natural language and feeding them into pre-trained generative models, enabling action inference without additional training [108, 109]. However, due to the limited scope of scenarios covered during pre-training, the models often struggle to generalize to the wide range of real-world decision-making tasks. Consequently, more studies have focused on fine-tuning generative models with domain-specific data [67,71–73,110, 111], or integrating them with additional networks [112–114] to improve the generalizability.

Task planning. Due to the limited capability of generative models in understanding low-level controls for many tasks, combined with the substantial resources and costs required for domain-specific fine-tuning, directly employing pre-trained generative models for high-level task planning presents a simpler and more widely applicable approach. Leveraging the reasoning [115] and reflection capabilities [68, 116–118] inherent, pre-trained generative

models can effectively make goal-oriented plans through iterative reasoning and reflection processes upon inputs such as environmental descriptions, states and objectives [49,119,120], as shown in Figure 9(b). For example, given a high-level task like "pour a glass of water and place it on the table", an LLM can decompose the task into a clear sequence of actionable steps: (S1) pick up the cup from the table, (S2) locate the water dispenser, (S3) align the cup with the dispenser's spout, (S4) activate the dispenser and wait until the cup is filled, and (S5) place the filled cup back on the table. Meanwhile, safety in embodied AI planning has garnered increasing attention in recent years [121,122].

Perception. Beyond decision-making, the strong expressive capabilities of generative models also make them effective in assisting embodied agents with environmental perception. For instance, HPT [123] utilizes the powerful sequence processing ability of Transformer architectures to fuse sensory data collected across multiple timesteps from various sources like images, IMU sensors, and radar, generating efficient observational representations, as shown in Figure 9(c). Similarly, numerous methods leverage Transformer, diffusion Model, and Mamba architectures for perceptual fusion [124–129]. Additionally, leveraging pre-trained VLMs for semantic understanding of multimodal perception data provides another effective approach to support embodied agents in perception tasks [64].

Reward design. To address the challenge of designing high-quality reward functions in complex real-world scenarios, utilizing pre-trained generative models for reward design has also become an effective approach to facilitate embodied agents in learning policies beyond perception. Reward design methods can be broadly categorized into reward signal generation and reward function generation. In reward signal generation methods, generative models receive state, action, and goal information at each decision step, producing reward signals in real-time [130,131]. Although this method offers flexibility and allows for dynamic adaptation to environmental changes, it incurs substantial computational costs due to frequent calls to generative models during training. In contrast, reward function generation methods only require a few pre-training calls to generative models before training to produce a reward function based on task descriptions and goals [132–136], as shown in Figure 9(d). During training, the reward function rather than the generative model is employed at each step, significantly reducing computational overhead.

Data-efficient learning. Furthermore, since interactions in real physical environments are costly and errors in robotic manipulation can lead to physical damage or system failures, employing generative models for data generation to enhance sample efficiency is another crucial approach in assisting policy learning [137–139]. Learning world models based on generative model architectures like Transformer and diffusion Models, and subsequently generating data via these world models, represents a mainstream method to boost sample efficiency [140–146], as shown in Figure 9(e). In higher-level language-abstracted environments, some methods directly use pre-trained generative models as world models for future trajectory predictions [147]. Another class of approaches, rather than learning explicit world models, utilizes generative models to create diverse simulation environments, generating additional data through interactions between agents and simulated environments [148–151].

3.4 Benchmarks

Although embodied agents are designed for interactive tasks in the real world, benchmarks still play a crucial role in standardizing performance measurement, guiding research focus, and reducing real-world testing costs. In recent years, there has been a surge of progress in the development of embodied AI benchmarks. We summarize several representative examples in Table 2. The listed benchmarks include the following.

- ALFRED [152] is a benchmark for evaluating embodied agents' ability to ground and execute free-form natural language instructions through physical interactions. Built on the AI2-THOR simulator, it includes 120 visually and functionally diverse household scenes, where agents are tasked with completing goals involving navigation, object manipulation, and irreversible state changes. These tasks span seven categories of household activities and involve 58 distinct object types. To support learning and evaluation, the dataset provides 25743 human-authored instructions paired with 8055 expert demonstrations. Agents perceive the environment from egocentric RGB-D inputs and act using a predefined set of 13 discrete low-level actions. Interaction targets are defined through pixel-wise masks, which guide object-level behaviors. Task performance is evaluated primarily based on goal completion, and the benchmark includes all code, annotations, and environment assets necessary to support reproducible research.
- RoboTHOR [153] is a sim-to-real embodied benchmark that provides paired simulated and physical environments designed for consistent cross-domain evaluations. It includes 75 training and validation scenes, along with 24 held-out test scenes (14 for test-dev and 10 for test-standard), which were all constructed from a modular asset library that supports flexible reconfiguration and expansion. The agents in RoboTHOR interact with the environment via a unified AI2-THOR API, and the policies can be remotely deployed on real LoCoBot robots whose noisy dynamics are closely matched to those in the simulations. RoboTHOR is centered on semantic navigation tasks

Table 2 Benchmarks for testing single-agent embodied AI.

	Year	Category			Input				View		Data			Agent type				
		Perception	Planning	Control	Image	Point cloud	Language	Proprioception	Global	Local	Real	Sim	Hybrid	Humanoid	Quadruped	UAV/UGV/USV	Manipulator	No certain type
ALFRED [152]	2020	✓	✓	✓	✓		✓			✓		✓					✓	
RoboTHOR [153]	2020	✓	✓	✓	✓					✓			✓				✓	
RobustNav [154]	2021	✓	✓	✓	✓			✓		✓		✓				✓		
Behavior [155]	2021	✓	✓	✓	✓		✓	✓		✓		✓		✓		✓	✓	
ManiSkill2 [156]	2023		✓	✓	✓	✓		✓		✓		✓					✓	
EgoCOT [157]	2023	✓	✓	✓	✓		✓			✓			✓				✓	
HumanoidBench [158]	2024	✓	✓	✓	✓			✓		✓		✓		✓			✓	
Open X-Embodiment [159]	2024	✓	✓	✓	✓	✓	✓			✓	✓					✓	✓	
OpenEQA [160]	2024	✓	✓	✓	✓	✓	✓			✓			✓			✓		
Embodied-Agent-Interface [161]	2024		✓	✓	✓		✓			✓		✓		✓				
ReALFRED [162]	2024	✓	✓	✓	✓		✓			✓		✓				✓	✓	
PhysBench [163]	2025	✓			✓		✓		✓				✓					✓

with performance evaluated using success rate, SPL and path efficiency in both virtual and physical settings. To promote reproducible research, all code, assets and remote-access infrastructure are publicly released.

• RobustNav [154] is a benchmark for evaluating the robustness of embodied navigation agents under realistic environmental and sensor corruptions. It extends standard PointGoal and ObjectGoal navigation tasks by introducing seven visual corruptions (e.g., motion blur, camera crack, low lighting) and three dynamics corruptions (e.g., motion bias, drift, motor failure) across 15 validation scenes. Agents are evaluated based on Success Rate and SPL, both before and after a fixed unsupervised "calibration budget." The benchmark highlights significant performance degradation under corruption, emphasizing the importance of robust perception, multimodal sensing, and adaptive policy learning.

• Behavior [155] is a benchmark for evaluating embodied AI agents in everyday household activities within virtual, interactive environments. It defines 100 realistic, diverse, and complex chores using a predicate logic-based language, and supports infinite scene-agnostic instantiations within iGibson 2.0. The benchmark includes 500 human VR demonstrations and provides evaluation metrics such as success score, task efficiency, and human-centric performance. All code, task definitions, and data are publicly available to facilitate reproducible research in embodied AI.

• ManiSkill2 [156] is a unified benchmark for generalizable robotic manipulation, comprising 20 task families that span stationary and mobile platforms, single- and dual-arm setups, and both rigid- and soft-body interactions. It includes over 2000 object models and more than 4 million demonstration frames. The suite supports point-cloud, RGB-D, and state-based inputs, with control at both the joint and end-effector levels. Built on the SAPIEN engine [164] and interfaced via OpenAI Gym [165], ManiSkill2 features asynchronous rendering and a warp MPM-based soft-body simulator for high-throughput data collection. An open-source codebase and online challenge platform support reproducible evaluation across perception-planning-execution pipelines, RL, and IL methods.

• EgoCOT and EgoVQA [157] provide 2900 hours of egocentric human-object interaction video clips accompanied by 3.85 million machine-generated, semantics-filtered, and human-verified chain-of-thought planning instructions. Complementarily, EgoVQA offers 200 million video-question-answer pairs automatically generated and CLIP-filtered from the Ego4D corpus. Together, these two datasets establish a unified, large-scale benchmark for end-to-end evaluation of embodied planning, control, and visual question answering in realistic first-person settings.

• HumanoidBench [158] is a high-dimensional simulation benchmark for humanoid robots, featuring 27 tasks including 15 whole-body manipulation tasks (e.g., truck unloading, basketball catch-and-throw) and 12 locomotion tasks (e.g., running, maze navigation). The benchmark uses a Unitree H1 humanoid equipped with dual Shadow Hands, yielding a 61-dimensional action space and a 151-dimensional observation space. Built on MuJoCo [166], it enables standardized evaluation of both flat and hierarchical RL methods in terms of success, efficiency, and robustness. Its open-source codebase and unified interfaces support reproducible research in complex humanoid control and learning.

• Open X-Embodiment [159] aggregates over one million real-world robot trajectories collected from 22 distinct

robot embodiments, spanning a total of 60 datasets. All trajectories are converted into a unified RLDS format [167], ensuring consistency and interoperability across platforms. To support broad adoption and reproducible research, the benchmark provides standardized APIs, a suite of open-source tools, and pretrained RT-X model checkpoints, enabling efficient training, evaluation, and comparison of generalist robotic policies across diverse embodiments and task domains.

• OpenEQA [160] is the first open-vocabulary benchmark for embodied question answering, featuring over 1600 human-authored questions grounded in more than 180 real-world scanned environments. It supports two evaluation paradigms—episodic memory and active exploration—where agents perceive the environment through RGB-D inputs and must reason about spatial, semantic, or temporal cues. Agent performance is assessed using an automated LLM-Match protocol, which enables consistent and reproducible evaluation of open-ended, language-driven understanding in complex 3D environments.

• Embodied-Agent-Interface [161] is a standardized benchmark for evaluating LLM-driven decision-making in embodied environments. It formalizes task goals using linear temporal logic and structures the decision-making process into four core modules: goal interpretation, sub-goal decomposition, action sequencing, and transition modeling. Each module is evaluated using fine-grained metrics such as grammar correctness, hallucination frequency, step completeness, and ordering accuracy. The benchmark is implemented on top of BEHAVIOR and VirtualHome, and incorporates automated sub-task checkers as well as PDDL-based planners to enable systematic evaluation and reproducible comparison across 18 LLMs in translating natural language instructions into executable plans.

• ReALFRED [162] is a photo-realistic benchmark for embodied instruction following, designed to bridge the gap between simulation and real-world complexity. It features 150 multi-room, 3D-scanned interactive environments, along with 30696 human-annotated natural language directives and planner-generated expert demonstrations spanning seven household task categories. Evaluation is based on standardized metrics for task success and execution efficiency, capturing both correctness and behavioral quality. To support reproducible research, ReALFRED provides open-source code and a complete dataset for studying vision-and-language-conditioned decision-making in realistic settings.

• PhysBench [163] is a large benchmark designed for evaluating VLMs on their understanding of physical concepts in real-world scenarios. It consists of 10002 multiple-choice instances that interleave video, image and text modalities, covering four core domains: physical object properties, object relationships, scene understanding and physical dynamics. These domains are further divided into 19 fine-grained sub-task types, enabling detailed assessment of model capabilities across diverse aspects of physical reasoning. The dataset of PhysBench is constructed from a combination of multiple resources including real-world video recordings, web-sourced media and high-fidelity physical simulations. To support fair and reproducible evaluation, PhysBench includes an online platform with withheld test answers, allowing standardized comparison across 75 representative VLMs.

4 Multi-agent embodied AI

Real-world embodied tasks often involve multiple agents or human-AI collaboration and competition, where dynamic interactions among agents in a shared environment give rise to emergent group-level behaviors that individuals cannot achieve alone. As a result, directly transferring methods designed for single-agent settings to MAS is typically inefficient. Therefore, MAS research primarily focuses on enabling effective collaboration. Compared to single-agent embodied intelligence, multi-agent settings introduce several unique challenges. First, the joint state-action space grows exponentially with the number of agents, making planning and learning significantly more complex. Second, partial observability is amplified, as each agent only has access to local perceptions in a shared and often dynamic environment. Third, the learning process becomes non-stationary due to concurrent updates from other agents. Finally, the problem of multi-agent credit assignment emerges, where it becomes difficult to attribute success or failure to individual agents. These challenges necessitate algorithmic advances tailored specifically for multi-agent embodied intelligence. Here, we review recent advances in multi-agent collaboration along the same three aspects as in the single-agent setting in Section 3: control and planning methods, learning-based methods and generative model-based methods, typical methods are listed in Table 3 [168–180].

4.1 Multi-agent control and planning

In MAS, control-based methods remain a fundamental approach for achieving high-precision, real-time decision-making under task constraints. Early methods model MAS as single agents and perform centralized control and planning [181–183] methods on them. However, these approaches face significant scalability challenges. To address this, some methods adopt distributed strategies that control each agent in the MAS independently [184], making

Table 3 Methods for building embodied AI in multi-agent settings.

	Year	Category			Input				View		Agent type					Implement		
		Perception	Planning	Control	Image	Point cloud	Language	Proprioception	Global	Local	Humanoid	Quadruped/Bipedal	UAV/UGV/USV	Manipulator	No certain type	Solver-based	NN-based	GM-based
ACE [168]	2023	✓	✓		✓					✓			✓				✓	
HetGPPO [169]	2023		✓	✓				✓		✓			✓				✓	
Self-play [170]	2023	✓	✓	✓	✓			✓	✓	✓			✓				✓	
TwoStep [171]	2024		✓				✓		✓						✓			✓
ZeroCAP [172]	2024	✓	✓		✓		✓		✓				✓					✓
COHERENT [173]	2024		✓		✓		✓		✓			✓	✓	✓				✓
LLaMAC [174]	2024	✓	✓	✓			✓		✓						✓			✓
FaGeL [175]	2024	✓	✓				✓			✓					✓			✓
MacNet [176]	2025		✓				✓		✓						✓			✓
CoELA [177]	2025	✓	✓	✓	✓	✓	✓	✓		✓	✓							✓
FISER [178]	2025		✓	✓			✓		✓						✓			✓
REMAC [179]	2025	✓	✓		✓		✓			✓			✓	✓				✓
LDPD [180]	2025		✓	✓				✓		✓			✓				✓	✓

them more suitable for large-scale MAS. Nevertheless, such fully decentralized methods often struggle with resolving conflicts between agents. To overcome these limitations, EMAPF [183] proposes a grouped multi-agent control framework. It dynamically clusters agents based on their spatial proximity, applying centralized control within each group while ensuring that inter-group control remains independent. This enables efficient coordination among large teams of aerial robots. Alternatively, Swarm-Formation [185] employs a decentralized sequential decision-making scheme: agents are pre-ordered and each agent plans its trajectory based on its predecessor's plan in turn, thereby eliminating inter-agent conflicts.

4.2 Learning for multi-agent interaction

Control-based methods in MAS, much like in single-agent scenarios, still face challenges such as high computational overhead and limited generalizability. As a result, learning-based approaches remain crucial in constructing multi-agent embodied AI. However, unlike single-agent settings, learning-based methods in MAS must additionally address unique challenges, including asynchronous decision-making, heterogeneous team composition and open multi-agent environments.

Asynchronous collaboration. In multi-agent embodied systems, challenges such as communication delays and hardware heterogeneity across agents often disrupt synchronized interactions and feedback from the real environment, making effective policy learning under asynchronous decision-making a significant challenge. To address this, ACE [168] introduces the concept of macro-actions, where a macro-action serves as a centralized goal for the entire MAS. Individual agents then make multiple asynchronous decisions based on this goal, with delayed feedback from the environment provided only after the macro-action is completed. To facilitate policy learning in this setup, ACE employs a macro-action-based MAPPO algorithm [168]. This approach has proven effective and has inspired several other studies tackling asynchronous decision-making challenges [186, 187].

Heterogeneous collaboration. Beyond differences in decision timing, another key distinction in embodied MAS is agent heterogeneity. This refers to the variations among agents in terms of perception capabilities, action spaces, task objectives, physical properties, communication abilities, and decision-making models. For instance, in a collaborative manufacturing scenario, autonomous vehicles might be responsible for transporting goods, while robotic arms handle sorting tasks. These two agent types differ significantly in their observation spaces, action spaces, and task goals, making them a prime example of heterogeneous agents. To tackle the challenges posed by such differences in observation and action spaces, approaches like HetGPPO [169] and COMAT [188] propose using separate observation and policy networks for different agent types. These networks are connected through graph-based communication, enabling effective information exchange. Another approach focuses on adapting the learning algorithm itself, such as by decomposing the advantage function across heterogeneous agents to facilitate more effective credit assignment [189].

Self-evolution in open environments. Unlike well-defined simulated scenarios, real-world embodied tasks typically take place in open environments, where key elements such as task objectives, environmental factors (e.g., states, actions, and reward functions), and collaboration patterns (e.g., teammates and opponents) evolve continuously and unpredictably [20]. To address these challenges in MARL, researchers have proposed methods such as robust training [190, 191] and continual coordination [192], which have yielded promising results. However, the unpredictable nature of open environments in embodied scenarios presents even greater difficulties, requiring specialized approaches beyond conventional MARL techniques. Recently, several innovative solutions leveraging the strong generalization capabilities of generative models have emerged. For example, when the number of collaborators dynamically changes, scalable architectures like graph neural networks (GNNs) and transformers can effectively encode interaction information [179]. By combining these architectures with distributed policy networks, agents can adapt seamlessly to fluctuations in team size, ensuring robust coordination [188]. Moreover, by incorporating self-play and other mechanisms for self-reflection and policy evolution, MAS can continuously improve collaborative performance in dynamic and open environments [170, 179].

4.3 Generative model-based multi-agent interaction

Although algorithms designed for asynchronous decision-making, heterogeneous agents and open environments have advanced learning-based methods in embodied MAS, challenges persist due to the diversity of cooperative behaviors and the partial or missing nature of observations, leading to low exploration efficiency and increased learning complexity. To address these issues, generative models have emerged as a powerful tool for enhancing decision-making in embodied MAS. They can introduce prior knowledge to facilitate explicit task allocation and leverage their strong information-processing capabilities to enable inter-agent communication and observation completion for distributed decision-making. Furthermore, MAS may not only involve autonomous agents but also humans. By harnessing the language understanding and generation capabilities of generative models, human-agent interaction and collaboration can be significantly improved, marking a unique and critical application in this domain. Lastly, the significantly larger exploration space in multi-agent settings makes the challenge of sample efficiency even more pronounced than in single-agent scenarios. Consequently, data augmentation techniques based on generative models are essential for improving data efficiency in multi-agent embodied AI.

Multi-agent task allocation. To address the challenges posed by the diverse collaborative behaviors in embodied MAS, an increasing number of studies have begun leveraging the prior knowledge and reasoning capabilities of pre-trained generative models to explicitly assign distinct tasks to different agents, significantly reducing the individual exploration spaces of each agent. For example, SMART-LLM [193] utilizes a pre-trained language model to first decompose a given task into multiple parallel sub-tasks and then group agents based on their capabilities. The sub-tasks are subsequently assigned to each group accordingly, enabling multi-agent decentralized task execution. For instance, different robotic vacuum cleaners are each allocated a distinct cleaning area. This paradigm of task decomposition and assignment has become a mainstream planning strategy in embodied MAS [171, 194, 195]. Building on this, a series of studies integrates task assignment and execution. These approaches first use generative models for task allocation, then continue using the same or similar models to execute the tasks, adjusting the allocation based on execution outcomes to form a more complete workflow [172, 173, 196, 197]. However, the aforementioned methods generally focus on task assignment based solely on generative reasoning, often overlooking the dependencies among sub-tasks. For instance, in the task "retrieve the wrench from a closed box", the sub-task "open the box" must be completed before "take the wrench out of the box". To better capture such dependencies, recent studies have focused on exploring the use of sub-task dependency graphs to enhance task allocation [198, 199]. In addition to explicit task decomposition, another line of work employs centralized generative models to produce global decisions. These are then implicitly distributed across agents by having each agent imitate the behavior generated by the centralized model, thus achieving implicit task assignment [180].

Multi-agent distributed decision-making. Using generative models for multi-agent embodied task decomposition and planning leverages the prior knowledge and reasoning capabilities acquired during pretraining to meet the demands of real-world collaborative tasks. However, centralized task planning and allocation can compromise the flexibility and scalability of cooperation, often requiring frequent recalls for adjustment. Therefore, exploring the coordination among multiple generative models is essential. Unlike independent methods in MARL that often fail in complex tasks, the outstanding perception and reasoning capabilities of generative models make it feasible to deploy multiple LLM-based agents that independently perform decision-making and policy evaluation effectively [200–206]. For example, Ref. [203] equipped each agent with a unique pre-trained generative model. These models leverage their strong information processing capabilities to communicate with other agents, either to complete missing observations or to request assistance, thus facilitating decentralized collaboration. However, chal-

lenges such as credit assignment and policy conflicts still hinder the effectiveness of fully distributed collaborative architectures. Therefore, similar to practices in MARL, some studies have begun to incorporate an additional centralized generative model to evaluate the decisions made by distributed generative models, enhancing their overall decision-making capabilities [200, 203]. Furthermore, by introducing a shared global LLM-based reflector to assess the contribution of each individual during collaboration, effective credit assignment in multi-LLM cooperation can also be achieved [202]. With appropriately designed system topologies and hierarchical collaboration frameworks, generative model-based distributed decision-making is capable of scaling to large-scale systems including up to thousands of agents [174, 176].

Beyond classic centralized or decentralized frameworks for embodied collaboration, generative models introduce a unique advantage: they allow negotiation among agents [203, 207, 208]. Unlike traditional collaboration pipelines where agents plan once every step, multiple LLMs are capable of engaging in iterative negotiation rounds during each planning phase, aiming to collectively identify an optimal course of action [208]. Beyond conventional coordination, multi-LLM negotiation enables embodied MAS to dynamically select and evolve their membership, organizational frameworks and leadership roles, thereby enhancing their capacity to handle complex tasks in open environments [203, 207].

Human-AI coordination. Efficient collaboration between humans and embodied agents has long been a critical objective of research [209]. Human-AI coordination [210], closely related to research fields such as human-AI interaction (HAI) [211] and human-robot interaction (HRI) [212], focuses on enhancing teamwork between humans and robots to effectively achieve complex tasks. Cooperative MARL, recognized for its robust problem-solving capabilities, provides promising avenues for improving human-AI collaboration across diverse user groups. However, traditional RL approaches often fall short in fully capturing the intricacies and variability inherent in human behaviors. Recently, with the rise of multimodal large models such as LLMs which inherently involve human-in-the-loop training processes, researchers have begun leveraging the extensive knowledge embedded in them to design sophisticated and adaptive strategies for human-AI collaboration [175, 177, 203, 213–218].

In scenarios including multiple LLM-based agents, collaboration between humans and agents can be partially facilitated by substituting human participants for communication partners or team leaders [177, 203, 213]. Nevertheless, this approach often under-utilizes the strengths that humans bring to the collaboration process. Given humans' superior capacities for understanding, reasoning, and adapting, recent methods have emphasized active human involvement to address limitations inherent in LLM-based agents [214]. For example, during navigation tasks, agents encountering uncertainty can proactively seek missing perceptual or decisional information from humans through natural language queries, allowing humans to offer guidance without directly engaging in detailed task planning or execution [219]. Additionally, LLM-based agents can proactively infer human intentions based on verbal and behavioral cues, facilitating more intuitive, flexible, and effective human-AI collaboration without explicit communication requests [178]. Furthermore, these agents can autonomously adapt and refine their collaborative behaviors over time, evolving through interactions without the need for explicit human instructions, thereby supporting sustained and effective long-term human-AI collaboration [175].

Data-efficient multi-agent learning. Due to the high sample efficiency of model-based RL approaches, applying world models to enhance multi-agent collaborative learning has always been a significant topic. In single-agent settings, world models serve as trajectory generators that provide data essential for policy learning. This involves predicting the future evolution of both the external environment and the agent's internal states based on past observations and actions, as well as estimating the potential rewards and state values associated with future trajectories. The policy is trained to maximize the predicted returns of imagined trajectories generated by the world model. To support this process, world models typically consist of a set of modular components for representation learning, dynamics prediction, and value estimation. These components work together to simulate plausible future rollouts without interacting with the real environment. A typical architecture for world model-augmented policy learning is as follows:

$$h_t = f_{\mathrm{ENC}}(s_t), \tag{11a}$$

$$\tilde{s}_t = f_{\mathrm{DEC}}(h_t), \tag{11b}$$

$$a_t = \pi(h_t), \tag{11c}$$

$$\hat{h}_{t+1} = f_{\mathrm{DYN}}(h_t, a_t), \tag{11d}$$

$$\hat{r}_t = f_{\mathrm{REW}}(h_t, a_t), \tag{11e}$$

$$\hat{Q}_t = f_{\mathrm{VAL}}(h_t, a_t), \tag{11f}$$

where f describes the components of the world model, s is the observed state, h is the latent state used to capture

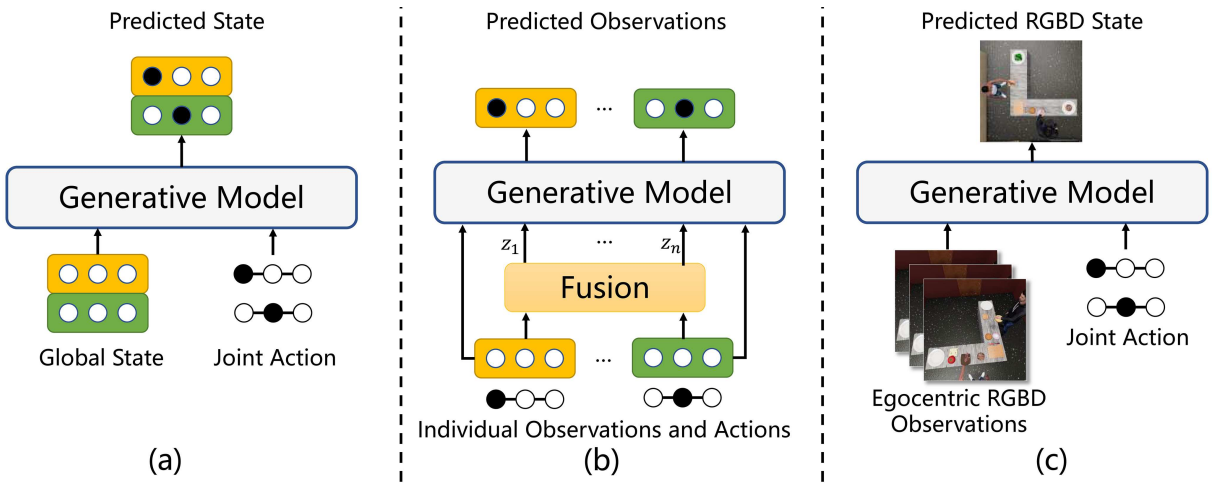


Figure 10 (Color online) An illustration of different multi-agent world model structures.

the dynamics in transitions and rewards, a is the action chosen by the policy π, $\hat{r}$ is the predicted reward and $\hat{Q}$ is the predicted state-action value. Starting from any collected observation s_t, the world model can iteratively interact with the policy to generate actions and update the latent state h forward in time, thereby producing data for policy learning without requiring interactions with the environment. During the execution phase, the learned world model can also function as a planner by selecting promising future trajectories for the agents.

However, modeling the interactions between agents and inferring the global state from local observations remains a key difficulty in leveraging world models to improve multi-agent collaboration efficiency. Early work attempted to model the dynamics of two collaborative robots simultaneously, achieving some success [220], as shown in Figure 10(a). However, as the number of collaborative agents increases, this approach quickly becomes inefficient. Through the use of generative models such as VAE and Transformers for local observation fusion, decoupling of global and local modeling and auto-regressive trajectory prediction, it is now possible to model the dynamic collaboration of MAS using world models [221–224], as shown in Figure 10(b). One approach addresses these challenges in the context of urban traffic by constructing a generative world model that simulates large-scale heterogeneous agents under physical and social constraints, using destination- and personality-conditioned dynamics to promote realistic and safe multi-agent behaviors [225]. Complementarily, another line of work focuses on physically embodied agents operating with egocentric RGBD observations, introducing a diffusion-based reconstruction process to establish a shared representation of the environment and incorporating intention inference to facilitate implicit coordination without direct communication [226], as shown in Figure 10(c).

4.4 Benchmarks

In contrast to the rapid advances observed in single-agent embodied intelligence, progress in multi-agent embodied intelligence has been relatively limited, especially in the development of standardized benchmarks. Existing benchmarks in this area frequently target highly specialized tasks or narrowly-defined scenarios, restricting their broader applicability and generalizability. To address this issue, this study provides a comprehensive and systematic review of current benchmarks for multi-agent embodied intelligence (see Table 4 [227–246]). Through detailed analyses and structured comparisons, we aim to offer researchers a clear overview of the state-of-the-art, highlight critical gaps in existing benchmarks, and facilitate the development of more robust, versatile, and widely applicable evaluation frameworks to support future research in this domain.

• Grid-MAPF [227] is a comprehensive benchmark for MAPF featuring 24 diverse grid maps that cover city layouts, game environments and warehouse-style grids, with 25 scenarios provided for each map. It systematically generates source-target pairs using random, clustered or designated methods and supports progressive agent scaling to test solver performance under increasing difficulty. Grid-MAPF formalizes various conflict types (e.g., vertex, edge, swapping) and supports multiple objectives such as makespan and sum of costs, enabling standardized and reproducible comparisons across MAPF algorithms.

• FurnMove [228] is a simulated benchmark designed to evaluate interactive furniture rearrangement, requiring agents to perceive, navigate, and manipulate multiple objects to achieve a target room layout. Built upon Habitat 2.0, it offers diverse 3D environments with varying furniture types and spatial configurations. Tasks are framed as

Table 4 Benchmarks for testing multi-agent embodied AI.

	Year	Category			Input				View		Data			Agent type				
		Perception	Planning	Control	Image	Point cloud	Language	Proprioception	Global	Local	Real	Sim	Hybrid	Humanoid/Human	Quadruped	UAV/UGV/USV	Manipulator	No certain type
Grid-MAPF [227]	2019		✓					✓	✓			✓						✓
FurnMove [228]	2020	✓	✓		✓		✓			✓		✓				✓	✓	
SMARTS [229]	2021	✓	✓	✓	✓			✓		✓		✓				✓		
Megaverse [230]	2021	✓	✓	✓	✓					✓		✓						✓
V2X-Sim [231]	2022	✓			✓	✓		✓		✓		✓				✓		
Bi-DexHands [232]	2022	✓	✓	✓	✓	✓		✓	✓	✓		✓					✓	
MRP-Bench [233]	2023	✓	✓	✓		✓		✓	✓	✓		✓				✓		
LEMMA [234]	2023	✓	✓	✓	✓		✓			✓		✓					✓	
CHAIC [235]	2024	✓	✓	✓	✓			✓		✓		✓						✓
RoCo [236]	2024	✓	✓	✓	✓		✓	✓	✓	✓			✓	✓		✓	✓	
MAP-THOR [237]	2024	✓	✓	✓	✓		✓	✓		✓		✓						✓
REMROC [238]	2024	✓	✓	✓		✓		✓		✓		✓				✓		
T2E [239]	2024	✓	✓	✓				✓		✓		✓				✓		
SocialGym 2.0 [240]	2024		✓	✓				✓		✓		✓				✓		
Cambridge RoboMaster [241]	2024	✓	✓	✓	✓			✓		✓			✓			✓		
S3E [242]	2024	✓			✓	✓		✓		✓	✓					✓		
PARTNR [243]	2024	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			✓			
EmbodiedBench [244]	2025	✓	✓	✓	✓		✓	✓		✓		✓				✓	✓	
MRMG-Planning [245]	2025		✓					✓	✓			✓				✓	✓	
CARIC [246]	2025	✓	✓	✓	✓	✓		✓		✓		✓				✓		

multi-step, object-centric transformations toward a predefined goal state. The benchmark defines structured scenarios with different levels of difficulty and introduces comprehensive evaluation metrics, including layout accuracy, motion efficiency, and collision count, making it a robust testbed for embodied planning and control in visually complex environments.

• SMARTS [229] is a scalable simulation platform and benchmark suite specifically designed for MARL in autonomous driving. It enables diverse and realistic interaction scenarios through a modular architecture, incorporating heterogeneous social agents from an extensible "Social Agent Zoo." SMARTS provides fine-grained simulation control via "bubbles" for localized agent interactions, supports distributed multi-agent training, and maintains compatibility with mainstream MARL libraries. The benchmark includes challenging driving tasks such as double merges and unprotected intersections, and offers a comprehensive set of evaluation metrics covering safety, agility, cooperation, and game-theoretic behavior, establishing it as a powerful testbed for complex MARL-based driving research.

• Megaverse [230] is an eight-task embodied AI benchmark. It features procedurally generated 3D environments designed to evaluate a wide range of agent capabilities, including maze navigation, obstacle traversal, exploration, puzzle solving, object manipulation and memory-based reasoning. The platform achieves high-throughput simulation with over 1 M observations per second through GPU-accelerated batched execution. It also supports multi-agent teams and enables infinite random instantiations of tasks, promoting generalization and robustness. As an open-source benchmark, Megaverse facilitates reproducible evaluation of embodied AI methods in navigation, exploration, planning, and manipulation.

• V2X-Sim [231] is a comprehensive benchmark for multi-agent collaborative perception in autonomous driving, supporting realistic vehicle-to-vehicle and vehicle-to-infrastructure communication scenarios. Built upon CARLA-SUMO co-simulation, it provides multimodal sensor data including RGB, LiDAR, GPS and semantic labels from both vehicles and roadside units. V2X-Sim evaluates collaborative 3D object detection, tracking and bird eye view semantic segmentation under various fusion strategies. With rich annotations, spatial diversity and support for intermediate feature sharing, V2X-Sim offers a rigorous testbed for assessing perception robustness under occlusion, sensor noise and communication limitations.

• Bi-DexHands [232] is a unified benchmark built on Isaac Gym [247] that achieves over 30000 FPS on a single

NVIDIA RTX 3090 by running thousands of environments in parallel, enabling highly efficient RL. It includes twenty bi-manual manipulation tasks inspired by the Fine Motor Sub-test, supporting single-agent, multi-agent, offline, multi-task and meta-RL settings within a coherent benchmark. Unlike prior benchmarks, Bi-DexHands treats each joint, finger and hand as a truly heterogeneous agent, and leverages a diverse set of everyday objects from the YCB [248] and SAPIEN datasets to evaluate task generalization. The platform also provides rich multimodal observations including contact forces, RGB images, RGB-D images and point clouds, and aligns task difficulty with human motor development milestones, enabling direct comparisons between robotic performance and infant dexterity.

• MRP-Bench [233] is an open-source benchmark for evaluating full-stack multi-robot navigation systems, integrating both centralized and decentralized MAPF algorithms as well as task assignment methods within realistic 3D environments from intralogistics and household domains. Built upon ROS2 [249], Gazebo [250], Nav2 package and the robotics middle-ware framework, it supports end-to-end evaluation of planning and coordination methods. MRP-Bench offers modular interfaces, configurable scenarios and standardized evaluation metrics including success rate, planning time, makespan and execution quality, enabling direct comparisons of multi-robot solutions under cluttered, dynamic and complex conditions.

• LEMMA [234] is a language-conditioned benchmark for multi-robot manipulation, consisting of eight procedurally generated, long-horizon tabletop tasks. Each involved task is designed to require true collaboration, as successful completion depends on two heterogeneous robot arms operating under minimal high-level or crowd-sourced natural language instructions. To support learning and evaluation, the benchmark provides 6400 expert demonstrations paired with corresponding language commands, along with rich multimodal observations such as RGB-D images and fused point clouds. The tasks are structured to enforce strong temporal dependencies and require exclusive multi-agent participation, reflecting the complexity of real-world collaborative manipulation. Evaluation is based on standardized success-rate metrics under a strict 100-second time constraint, enabling comprehensive assessment of agents' capabilities in language-guided planning, task allocation, and fine-grained coordination.

• CHAIC [235] introduces the first large-scale benchmark for embodied social intelligence with a specific focus on accessibility. It simulates interactions between helper agents and four types of physically constrained human partners, including wheelchair users, children, frail adults, and bicyclists. These interactions are carried out across eight long-horizon tasks set in both indoor and outdoor environments, many of which involve emergency scenarios. To successfully assist their partners, agents must infer individual goals and physical limitations based on egocentric RGB-D observations, and generate user-adaptive cooperative plans grounded in realistic physics. The benchmark evaluates performance using metrics such as transport success rate, efficiency improvement, and goal inference accuracy, enabling rigorous assessment of socially aware embodied collaboration.

• RoCo [236] is a benchmark consisting of six tasks for evaluating multi-robot manipulation. It is specifically designed to probe key dimensions of collaboration, including task decomposition into parallel or sequential subtasks, symmetry of observations between agents, and varying degrees of workspace overlap. The environment is built on MuJoCo and is accompanied by a text-only dataset. Tasks in RoCo involve everyday household objects and require agents to demonstrate zero-shot adaptation, feedback-driven re-planning and language-grounded task reasoning. The evaluations in RoCo are based on metrics including success rate, environment step efficiency and the number of re-planning attempts, allowing rigorous assessment of both high-level coordination and low-level motion planning performance.

• MAP-THOR [237] is a benchmark consisting of 45 long-horizon, language-instructed multi-agent tasks situated in the AI2-THOR simulator under partial observability. To evaluate spatial generalization, each task is instantiated across five distinct floor plans. The benchmark also categorizes tasks into four levels according to the explicitness of item types, quantities, and target locations. These levels span a spectrum from fully specified instructions to entirely implicit goals, progressively increasing the demands on planner robustness in the face of linguistic ambiguity. To support consistent and comprehensive evaluation, the benchmark adopts standardized metrics such as success rate, transport rate, coverage, balance, and average decision steps. Together, these metrics enable rigorous end-to-end comparisons of multi-agent planning approaches without relying on privileged simulator information.

• REMROC [238] is a modular, open-source benchmark designed to evaluate multi-robot coordination algorithms in realistic and unstructured environments that are shared with humans. It is built on ROS2, Gazebo Ignition and Navigation2, and supports heterogeneous robot platforms, configurable environments with varying human densities, and integrated human motion models. To enable systematic assessment, the benchmark records a variety of performance metrics, including time-to-goal and path efficiency, which allow researchers to quantify the impact of human presence on coordination effectiveness.

• T2E [239] is a 2D simulation benchmark where heterogeneous captor and target robots must coordinate to trap opponents, rather than simply pursue them. This is achieved by combining inter-agent cooperation with

environmental constraints. The trapping process is formalized using the absolutely safe zone (ASZ) metric, which measures the degree of spatial confinement. The benchmark includes open-source environments based on real-world maps and provides tailored MARL baselines through a unified TrapNet architecture. It supports both fully cooperative and competitive evaluation settings. Standardized metrics, including completion time, success rate, path length, and ASZ area change, are used to rigorously assess collaborative behaviors in complex spatial layouts.

• SocialGym 2.0 [240] is a multi-agent social navigation benchmark for training and evaluating autonomous robots with individual kino-dynamic constraints. It supports a variety of environments including both open spaces and geometrically constrained areas such as doorways, hallways, intersections and roundabouts. The simulator integrates the PettingZoo [251] with ROS messaging, allowing for fully customizable observation and reward functions. SocialGym2.0 includes a set of social mini-games along with realistic building layouts. It defines evaluation metrics such as success rate, collision rate and stopping time, which are used to rigorously assess policy performance in diverse social interaction scenarios.

• Cambridge RoboMaster [241] offers an agile and low-cost research testbed for MAS by retrofitting DJI RoboMaster S1 platforms with advanced onboard compute such as NVIDIA Jetson Orin, custom ROS2 drivers, and dual-network communication capabilities using both CAN and WiFi. The platform supports seamless sim-to-real transfer through integration with the VMAS framework [252], allowing zero-shot deployment of both traditional model-based controllers and decentralized MARL policies driven by GNNs. This hardware-software integration enables real-world experimentation under controlled yet realistic conditions. The testbed supports a variety of embodied multi-agent tasks, including trajectory tracking, visual SLAM, and cooperative navigation, thereby facilitating reproducible, scalable, and physically grounded evaluation of coordination strategies and policy generalization in embodied systems.

• S3E [242] is a large-scale benchmark for collaborative SLAM (C-SLAM). It features synchronized 360° LiDAR, high-resolution stereo imagery, high-frequency IMU and novel UWB ranging data collected from three mobile platforms. The dataset covers 18 indoor and outdoor sequences and totals over 263 GB. To support systematic evaluation, S3E defines four representative trajectory patterns: concentric circles, intersecting circles, intersection curve and rays. These patterns test the ability of C-SLAM methods to handle both intra- and inter-loop closures under different levels of spatial overlap.

• PARTNR [243] is a large-scale benchmark for embodied human-AI collaboration driven by natural language. It includes 100000 procedurally generated tasks distributed across 60 diverse multi-room houses containing 5819 unique objects. All the included tasks are divided into four types: constraint-free, spatial, temporal and heterogeneous, which are designed to evaluate long-horizon planning and inter-agent coordination under realistic physical conditions. Each instruction is paired with a simulation-grounded evaluation function that is automatically generated, enabling rigorous assessment of task completion. The benchmark is built using a semi-automated pipeline that integrates LLM-based task generation, evaluation synthesis, in-the-loop simulation filtering, and human validation. This design enables PARTNR to reach an unprecedented scale and diversity. It also highlights the performance gap between current LLM-based planners and human collaborators in complex multi-agent embodied tasks.

• EmbodiedBench [244] provides a comprehensive suite of 1128 testing instances across four embodied environments. These include household scenarios (EB-ALFRED and EB-Habitat) and low-level control domains (EB-Navigation and EB-Manipulation). The tasks span both high- and low-level action hierarchies and are organized into six capability-oriented subsets: base, common sense, complex instruction, spatial awareness, visual perception and long-horizon reasoning. The benchmark supports fine-grained, multimodal evaluation of 19 state-of-the-art proprietary and open-source multimodal large language models (MLLMs), enabling systematic assessment across a wide range of embodied AI challenges.

• MRMG-Planning [245] formalizes the problem of multimodal, multi-robot, multi-goal path planning in continuous spaces. The benchmark includes 21 diverse scenarios, ranging from simple validation tasks to complex cases involving up to five heterogeneous robots and 74 sequential goals. MRMG-Planning supports dynamic environment changes by modeling mode transitions that modify both robot assignments and scene geometry, such as during object handovers. Each scenario is instantiated in a composite configuration space and paired with two baseline planners based on RRT* and PRM*, both of which are asymptotically optimal and probabilistically complete. This design enables rigorous evaluation of end-to-end planning performance under varying robot counts, goal sequences and cost formulations.

• CARIC [246] provides a benchmark for heterogeneous multi-UAV inspection, in which LiDAR-equipped explorers and camera-only photographers must collaboratively map unknown structures using only bounding box constraints while capturing high-quality structural imagery. CARIC incorporates realistic limitations on kinematics, communication and sensing capabilities. It includes a variety of inspection scenarios such as buildings, cranes and aircraft, and defines rich evaluation metrics that decompose the inspection score into components for visibility,

motion blur and spatial resolution. These metrics enable systematic evaluation of task allocation and planning under tight constraints and support reproducible comparisons for real-world aerial inspection tasks demonstrated in CDC 2023 and IROS 2024.

5 Challenges and future work

Despite the rapid progress of embodied AI and the primary development of multi-agent embodied AI, it faces several challenges and presents exciting future directions.

Theory for complex embodied AI interaction. Building upon the theoretical foundations provided by Markov games, MARL has introduced various frameworks for effectively modeling cooperative relationships among agents in complex environments [19]. Methods such as value decomposition (e.g., VDN [253], QMIX [46]) have improved scalability by decomposing joint objectives into individual utilities, while counterfactual reasoning approaches like COMA [254] have significantly enhanced attribution accuracy for individual agent contributions. Advances in transfer learning and networked MARL have mitigated challenges posed by agent heterogeneity and environmental non-stationarity. Additionally, the integration of control and game theory has fostered effective coordination strategies, exemplified by consensus protocols and distributed task allocation methods [255]. However, embodied multi-agent intelligence introduces unique theoretical challenges, including asynchronous sensing, delayed actions, limited observability, communication constraints, and substantial heterogeneity, complicating both theoretical formulation and practical implementation. Furthermore, while large generative models, particularly LLMs, offer significant potential in facilitating planning and inter-agent communication, their theoretical characteristics such as stability, generalization and interpretability remain inadequately understood [256]. To address these limitations, future research should explore new theoretical paradigms specifically tailored to embodied MAS, leveraging approaches such as causal inference to uncover inter-agent dependencies, complex systems theory to understand emergent behaviors, and hierarchical, bio-inspired coordination frameworks adaptable to real-world complexity.

New algorithm design. Recent developments have solidified the role of MARL as a foundational framework for cooperative and competitive multi-agent scenarios. Notable algorithms, including QMIX [46] with its monotonic value decomposition and MAPPO [48], an adaptation of proximal policy optimization, have demonstrated compelling results in benchmark environments such as SMAC and Google research football (GRF) [20]. Nevertheless, these successes predominantly rely on the CTDE paradigm [257], assuming ideal conditions and unrestricted access to agent actions—conditions rarely achievable in embodied scenarios. Physical deployment introduces significant complexities such as sensor noise, limited actuation, delayed feedback, and partial observability. Moreover, existing multi-agent learning frameworks typically struggle with generalization beyond training distributions and have limited scalability when dealing with heterogeneous teams. Embodied multi-agent tasks involve multimodal sensory data, diverse capabilities, and dynamic interactions, necessitating alternative algorithmic frameworks beyond traditional paradigms such as CTDE, DTDE (decentralized training and execution), or CTCE (centralized training and execution). Promising future directions include hierarchical coordination structures, agent-grouping mechanisms, and the integration of structured priors or classical control theories to achieve scalable and robust performance in embodied MAS.

Effective and efficient learning. Most multi-agent learning methodologies have been developed in simulation or gaming environments, benefiting from low-cost, highly repeatable scenarios that allow extensive policy optimization through frequent interaction sampling. However, embodied multi-agent tasks introduce significantly greater complexity than their counterparts in simulated or game-based environments. In the real world, both the state and policy spaces become more intricate due to high-dimensional sensory inputs, continuous dynamics, and diverse task variations. Moreover, collecting interaction data is often expensive, time-consuming, and subject to physical constraints. These challenges are further amplified in real-world multi-agent contexts, where multiple autonomous agents must coordinate their behaviors under constraints such as considerable time, financial cost, and hardware wear, making joint exploration and sample-efficient learning substantially more difficult [21, 258]. Although recent advancements in single-agent embodied learning—such as extracting knowledge from large multimodal models like LLMs [259], employing world-model-based simulations for policy rollouts [260], and leveraging offline data-driven policy calibration [261]—have improved sample efficiency, these strategies often struggle in multi-agent contexts owing to complex interactions and non-stationary dynamics. To overcome these barriers, specialized multi-agent world models accurately simulating interactive dynamics, exploration strategies informed by structured prior knowledge, rapid initialization through IL, and meta-learning for improved generalization are necessary. Additionally, developing robust methodologies supporting effective multi-task and sim-to-real transfer is vital for real-world applicability.

Large generative models assisted learning. Recent advances in pre-trained large models, such as GPT-4,

PaLM, CLIP, SAM and Gemini, have significantly reshaped language, vision and multimodal learning landscapes by offering powerful capabilities in representation, perception, and reasoning [262]. These foundation models, trained on vast and diverse datasets, provide rich prior knowledge and robust cross-modal alignment, presenting promising avenues for enhancing embodied AI. Leveraging these models allows agents to attain deeper semantic understanding, superior generalization, and more adaptive interactions. However, directly deploying these foundation models in embodied multi-agent scenarios remains challenging [194], as they typically derive from static, single-agent contexts and lack inductive biases essential for dynamic interactions involving asynchronous communication, partial observability, tightly coupled policies, and non-stationarity. Furthermore, embodied MAS usually involves high-dimensional multimodal inputs, extensive action spaces and sparse feedback, limiting the adaptability and efficiency of the learning algorithms. Insights from recent single-agent embodied learning—such as few-shot adaptation, multimodal joint pretraining and simulation-driven data augmentation, provide valuable transferable knowledge [16] for multi-agent research. Future research should focus on developing scalable multi-agent pretraining paradigms, integrating foundation models with reinforcement learning and graph-based coordination methods, and enhancing sim-to-real transfer, thereby forming robust and generalizable frameworks for multi-agent embodied AI in complex, open-ended environments.

General multi-agent embodied AI framework. Recent advancements in multi-agent embodied AI have primarily emphasized solving specific tasks within constrained scenarios. While progress is notable, this focus significantly limits the generalizability and scalability of policies. Benchmarks such as SMAC and GRF have driven progress in MARL, yet existing methods continue to struggle with variable task objectives, environmental dynamics, and agent team compositions [20]. Single-agent generalist models like Gato [263] and RT-X [159] demonstrate the potential of unified architectures in broadly generalizing across vision, language, and control domains. Extending generalist principles to multi-agent settings introduces additional complexity from inherent non-stationarity, intricate agent interactions, and multiple equilibria. Moreover, current architectures often assume fixed team sizes and agent homogeneity, restricting scalability and adaptability. Recent innovations such as the multi-agent Transformer (MAT) [264] address scalability via permutation-invariant representations but substantial challenges remain, particularly in integrating multimodal knowledge (e.g., physical embodiment, social reasoning, linguistic interaction) into unified frameworks. Future advancements will require equilibrium-aware training methodologies, modular and scalable architectures, and robust multimodal representation learning techniques, enabling reliable coordination and generalization in dynamic multi-agent environments.

Adaptation to open environments. Unlike the stable and closed task settings commonly assumed in recent embodied AI research, open environments [265] more accurately capture the complexities of the real world, where dynamics are inherently uncertain, non-stationary, and unpredictable. This increased realism introduces significantly greater challenges for embodied MAS, as agents must operate under constantly changing conditions with limited prior knowledge. Nevertheless, addressing these challenges is crucial, as success in open environments provides more meaningful insights and practical value for real-world deployment. Agents operating within such environments often encounter unpredictable sensory inputs, shifting reward structures, evolving tasks and frequently changing agent populations. For instance, the environments may include noisy observations, delayed or disrupted feedback and task transitions ranging from navigation and exploration to cooperative or competitive interactions. Additionally, the shifting team compositions with teammates and adversaries changing unpredictably further complicate coordination. Human involvement also introduces extra layers of uncertainty due to diverse, less predictable behaviors. Such variability undermines the consistency between training and deployment environments, revealing the fragility of policies [20]. Thus, achieving effective performance in open environments requires agents with robust policies resilient to sensory and environmental disturbances, continual learning capabilities that integrate new experiences without compromising previously acquired skills and contextual reasoning abilities for dynamic interactions and coordination with novel agents. Methods enabling rapid strategy adaptation such as policy reuse, latent variable modeling and meta-learning have become indispensable for generalizing quickly to unseen circumstances. Although recent advances in robust MARL, lifelong learning, and context-aware decision-making are promising, critical challenges remain unsolved. Future research should prioritize accurate modeling of open-environment dynamics, anticipating heterogeneous agent and human behaviors, and enabling real-time algorithmic adaptation.

Evaluation and verification. Bridging the gap between simulation and reality remains a central challenge for advancing multi-agent embodied AI, particularly in the development of rigorous and comprehensive evaluation frameworks. While recent platforms, such as CHAIC [235] (see additional benchmarks in Table 4), have contributed valuable resources, they often exhibit limited representational complexity and lack ecological validity. Current benchmarks predominantly focus on single-modality tasks, neglecting essential multimodal signals (vision, language, audio, video) necessary for realistic perception and decision-making. Furthermore, benchmarks commonly involve small-scale, homogeneous MAS, inadequately representing the complexity encountered in heterogeneous

MAS comprising diverse embodiments such as drones, robotic arms, autonomous vehicles, and quadrupedal robots. Additionally, idealized assumptions, such as ignoring partial observability, communication latency, adversarial behaviors, and asynchronous execution, widen the sim-to-real gap, undermining policy robustness and transferability. To help bridge this gap, physical testbeds—such as robot soccer platforms (e.g., RoboCup [266]) and drone swarm arenas—are essential for validating real-world feasibility and grounding embodied evaluation in physical interactions. At the same time, the lack of standardized and interpretable evaluation metrics remains an unresolved challenge, as existing measures (e.g., coordination scores, network-based metrics) often suffer from inconsistency and poor generalizability, making them inadequate for evaluating complex real-world embodied tasks. Subsequently, future research should aim to develop general-purpose, modular and scalable evaluation frameworks analogous to Gym, MuJoCo, or PyMARL [267], incorporating multimodal interactions, heterogeneous coordination, and reproducibility features. Therefore, developing unified benchmarks with clearly defined tasks, comprehensive evaluation criteria (robustness, scalability, energy efficiency, behavioral diversity), open-access leaderboards, and formal verification methods is imperative, especially for safety-critical applications.

Applications and implementations. Multi-agent embodied AI holds immense promise across a wide range of domains, including robotics, education, healthcare, military operations, interactive simulations, and smart-city infrastructures [268, 269]. In robotics, particularly industrial manufacturing, warehouse logistics, and autonomous driving, effective deployment demands robust real-time coordination, reliable collision avoidance, and adaptable hardware-software integration, thereby emphasizing the need for algorithmic generalizability. In educational settings, where virtual agent teams interact with learners, socially appropriate and pedagogically effective behaviors are required, placing high demands on natural language understanding and affective computing. In healthcare, the integration of robotic agents into clinical workflows highlights the importance of trustworthy decision-making, stringent privacy protection, and adherence to ethical standards. Military applications, such as coordinated drone swarms, further underscore these challenges, requiring policy robustness under adversarial conditions and human oversight to mitigate ethical and operational risks associated with autonomy. Interactive simulation and competitive gaming platforms such as OpenAI Five [270] and AlphaStar [271] have served as valuable testbeds, though their high computational demands and limited transferability to real-world settings pose ongoing limitations. In smart-city scenarios, including traffic control and energy-grid management, large-scale multi-agent coordination must be highly reliable, efficient, and safety-critical, directly impacting societal operations. Addressing these diverse yet interconnected challenges calls for the integration of domain-specific knowledge, rigorous validation pipelines, and comprehensive safety frameworks to enable the trustworthy and scalable deployment of embodied multi-agent systems in the real world.

Other future directions. While the previous sections highlight significant challenges within multi-agent embodied AI, several additional fundamental issues warrant deeper investigation. First, multimodal perception and collaborative learning remain underexplored, with current research typically focusing on isolated sensory modalities and rarely effectively integrating vision, audio and language inputs. Future studies should thus prioritize developing robust cross-modal fusion methods and efficient communication protocols to handle modality disparities, information inconsistencies and latency issues [272]. Second, social learning presents a promising pathway toward emergent collective AI. Although approaches such as IL and policy distillation enable agent populations to exhibit coordination surpassing individual capabilities, theoretical foundations and conditions necessary for stable emergent behaviors remain inadequately understood [273]. Lastly, as MAS are increasingly gaining autonomy, addressing safety and ethical considerations becomes critically important, especially within sensitive or high-stakes environments [274]. Explicit safety constraints, clear ethical guidelines and transparent explainability mechanisms must be developed to ensure fair, accountable and interpretable decision-making, thereby minimizing unintended risks arising from complex agent interactions and ensuring trustworthy AI deployments.

6 Conclusion

Embodied AI equips agents with the capability to sense, perceive and effectively interact with various entities in both cyberspace and the physical world, marking a significant step toward achieving artificial general intelligence. To provide a systematic understanding of this rapidly evolving field, this survey comprehensively reviews recent developments in embodied AI, bridging the gap from single-agent to multi-agent scenarios, and further discusses key challenges and promising future directions. Specifically, we first introduce foundational concepts essential to embodied AI. We then systematically examine critical methodologies, including classical control and planning approaches, learning-based techniques, and generative-model-based frameworks. Additionally, widely adopted benchmarks employed for evaluating embodied AI in both single-agent and multi-agent contexts are carefully discussed. Finally,

we highlight significant challenges and explore promising future directions within this emerging research domain. Through a comparative analysis of recent advances, this survey aims to offer researchers a clear and integrated perspective, thereby facilitating continued innovation in multi-agent embodied AI.

Acknowledgements The work was supported by National Natural Science Foundation of China (Grant Nos. 62495090, 62495095, U23B2059).

References

1 Pfeifer R, Iida F. Embodied artificial intelligence: trends and challenges. In: Embodied Artificial Intelligence. Berlin: Springer, 2004. 1–26
2 Duan J, Yu S, Tan H L, et al. A survey of embodied AI: from simulators to research tasks. IEEE Trans Emerg Top Comput Intell, 2022, 6: 230–244
3 Turing A M. I.—Computing machinery and intelligence. Mind, 1950, 49: 433–460
4 Brooks R A. New approaches to robotics. Science, 1991, 253: 1227–1232
5 Wu D, Wei X, Chen G, et al. Generative multi-agent collaboration in embodied AI: a systematic review. 2025. ArXiv:2502.11518
6 Wang X, Chen G, Qian G, et al. Large-scale multi-modal pre-trained models: a comprehensive survey. Mach Intell Res, 2023, 20: 447–482
7 Liu Y, Chen W, Bai Y, et al. Aligning cyber space with physical world: a comprehensive survey on embodied AI. 2024. ArXiv:2407.06886
8 Wang Z, Yu B, Zhao J, et al. Karma: augmenting embodied AI agents with long-and-short term memory systems. In: Proceedings of 2025 IEEE International Conference on Robotics and Automation, 2025. 1–8
9 Ziliotto F, Campari T, Serafini L, et al. Tango: training-free embodied AI agents for open-world tasks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025. 24603–24613
10 Roy N, Posner I, Barfoot T, et al. From machine learning to robotics: challenges and opportunities for embodied intelligence. 2021. ArXiv:2110.15245
11 Newell A, Simon H A. Computer science as empirical inquiry: symbols and search. Commun ACM, 1976, 19: 113–126
12 Brooks R. A robust layered control system for a mobile robot. IEEE J Robot Automat, 1986, 2: 14–23
13 Zare M, Kebria P M, Khosravi A, et al. A survey of imitation learning: algorithms, recent developments, and challenges. IEEE Trans Cybern, 2024, 54: 7173–7186
14 Eppe M, Gumbsch C, Kerzel M, et al. Hierarchical principles of embodied reinforcement learning: a review. 2020. ArXiv:2012.10147
15 Xu Z, Wu K, Wen J, et al. A survey on robotics with foundation models: Toward embodied AI. 2024. ArXiv:2402.02385
16 Ma Y, Song Z, Zhuang Y, et al. A survey on vision-language-action models for embodied AI. 2024. ArXiv:2405.14093
17 Dorri A, Kanhere S S, Jurdak R. Multi-agent systems: a survey. IEEE Access, 2018, 6: 28573–28593
18 Xu J, Sun Q, Han Q L, et al. When embodied AI meets industry 5.0: human-centered smart manufacturing. IEEE CAA J Autom Sin, 2025, 12: 485–501
19 Zhang K, Yang Z, Başar T. Multi-agent reinforcement learning: a selective overview of theories and algorithms. In: Handbook of Reinforcement Learning and Control. Cham: Springer, 2021. 321–384
20 Yuan L, Zhang Z, Li L, et al. A survey of progress on cooperative multi-agent reinforcement learning in open environment. 2023. ArXiv:2312.01058
21 Albrecht S V, Christianos F, Schäfer L. Multi-agent Reinforcement Learning: Foundations and Modern Approaches. Cambridge: MIT Press, 2024
22 Wang G, Liu W J, Li Y F, et al. Data-driven control of network systems: accounting for communication adaptivity and security. Sci China Inf Sci, 2026, 69: 121201
23 Oroojlooy A, Hajinezhad D. A review of cooperative multi-agent deep reinforcement learning. Appl Intell, 2023, 53: 13677–13722
24 Li Z, Wu W, Guo Y, et al. Embodied multi-agent systems: a review. IEEE CAA J Autom Sin, 2025, 12: 1095–1116
25 Tarafdar R. Embodied and multi-agent reinforcement learning: advances, challenges and opportunities. Int J Innovative Sci Res Tech, 2025, 10: 3183–3187
26 Bridgeman B, Tseng P. Embodied cognition and the perception-action link. Phys Life Rev, 2011, 8: 73–85
27 Zhao W X, Zhou K, Li J, et al. A survey of large language models. 2023. ArXiv:2303.18223
28 Guo T, Chen X, Wang Y, et al. Large language model based multi-agents: a survey of progress and challenges. 2024. ArXiv:2402.01680
29 Ross I M. A Primer on Pontryagin's Principle in Optimal Control. Springfield: Collegiate Publishers, 2015
30 Sutton R S, Barto A G. Reinforcement Learning: An Introduction. Cambridge: MIT Press, 2018
31 Lample G, Chaplot D S. Playing FPS games with deep reinforcement learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2017
32 Souchleris K, Sidiropoulos G K, Papakostas G A. Reinforcement learning in game industry—review, prospects and challenges. Appl Sci, 2023, 13: 2443
33 Singh B, Kumar R, Singh V P. Reinforcement learning in robotic applications: a comprehensive survey. Artif Intell Rev, 2022, 55: 945–990
34 Nian R, Liu J, Huang B. A review on reinforcement learning: Introduction and applications in industrial process control. Comput Chem Eng, 2020, 139: 106886
35 Yu C, Liu J, Nemati S, et al. Reinforcement learning in healthcare: a survey. ACM Comput Surv, 2021, 55: 1–36
36 Kiran B R, Sobh I, Talpaert V, et al. Deep reinforcement learning for autonomous driving: a survey. IEEE Trans Intell Transp Syst, 2022, 23: 4909–4926
37 Li Z, Wu W R, Wang J L, et al. Continuous advantage learning for minimum-time trajectory planning of autonomous vehicles. Sci China Inf Sci, 2024, 67: 172206
38 Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback. In: Proceedings of Advances in Neural Information Processing Systems, 2022
39 Havrilla A, Du Y, Raparthy S C, et al. Teaching large language models to reason with reinforcement learning. 2024. ArXiv:2403.04642
40 Cao H Y, Feng F, Huo J, et al. Causal action empowerment for efficient reinforcement learning in embodied agents. Sci China Inf Sci, 2025, 68: 150201
41 Zhou X H, Xiong J, Zhao H T, et al. Joint UAV trajectory and communication design with heterogeneous multi-agent reinforcement learning. Sci China Inf Sci, 2024, 67: 132302
42 Kang Y, Di J, Li M, et al. Autonomous multi-drone racing method based on deep reinforcement learning. Sci China Inf Sci, 2024, 67: 180203
43 Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning. 2013. ArXiv:1312.5602
44 Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning. 2015. ArXiv:1509.02971
45 Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms. 2017. ArXiv:1707.06347
46 Rashid T, Samvelyan M, De Witt C S, et al. Monotonic value function factorisation for deep multi-agent reinforcement learning. J Mach Learn Res, 2020, 21: 1–51
47 Lowe R, Wu Y, Tamar A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments. In: Proceedings of Advances in Neural Information Processing Systems, 2017. 6382–6393
48 Yu C, Velu A, Vinitsky E, et al. The surprising effectiveness of PPO in cooperative multi-agent games. In: Proceedings of Advances in Neural Information Processing Systems, 2022. 24611–24624

49 ichter b, Brohan A, Chebotar Y, et al. Do as I can, not as I say: grounding language in robotic affordances. In: Proceedings of the 6th Conference on Robot Learning, 2023. 287–318
50 Pateria S, Subagdja B, Tan A, et al. Hierarchical reinforcement learning: a comprehensive survey. ACM Comput Surv, 2021, 54: 1–35
51 Song Y, Scaramuzza D. Policy search for model predictive control with application to agile drone flight. IEEE Trans Robot, 2022, 38: 2114–2130
52 Feng Z, Chen J, Xiao W, et al. Learning hybrid policies for MPC with application to drone flight in unknown dynamic environments. Un Sys, 2024, 12: 429–441
53 Duan K, Dong S, Fan Z, et al. Multimode trajectory tracking control of unmanned surface vehicles based on LSTM assisted model predictive control. Ocean Eng, 2025, 328: 121015
54 Yang X, Ji Z, Wu J, et al. Hierarchical reinforcement learning with universal policies for multistep robotic manipulation. IEEE Trans Neural Netw Learn Syst, 2022, 33: 4727–4741
55 Li J, Tang C, Tomizuka M, et al. Hierarchical planning through goal-conditioned offline reinforcement learning. IEEE Robot Autom Lett, 2022, 7: 10216–10223
56 Ho J, Ermon S. Generative adversarial imitation learning. In: Proceedings of Advances in Neural Information Processing Systems, 2016. 4572–4580
57 Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Proceedings of Advances in Neural Information Processing Systems, 2014. 2672–2680
58 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Proceedings of Advances in Neural Information Processing Systems, 2017
59 Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. In: Proceedings of Advances in Neural Information Processing Systems, 2020. 6840–6851
60 Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces. In: Proceedings of the First Conference on Language Modeling, 2024
61 Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Proceedings of Medical Image Computing and Computer-Assisted Intervention, 2015. 234–241
62 Lavavalle S. Rapidly-exploring random trees: a new tool for path planning. Research Report 9811, 1998
63 Wang Z, Zhou X, Xu C, et al. Geometrically constrained trajectory optimization for multicopters. IEEE Trans Robot, 2022, 38: 3259–3278
64 Zheng S, Liu J, Feng Y, et al. Steve-eye: equipping llm-based embodied agents with visual perception in open worlds. 2023. ArXiv:2310.13255
65 Fu J, Song Y, Wu Y, et al. Learning deep sensorimotor policies for vision-based autonomous drone racing. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2023. 5243–5250
66 Ze Y, Yan G, Wu Y H, et al. GNFactor: multi-task real robot learning with generalizable neural feature fields. In: Proceedings of the 7th Annual Conference on Robot Learning, 2023
67 Driess D, Xia F, Sajjadi M S M, et al. PaLM-e: an embodied multimodal language model. In: Proceedings of the 40th International Conference on Machine Learning, 2023. 8469–8488
68 Huang W, Xia F, Xiao T, et al. Inner monologue: embodied reasoning through planning with language models. In: Proceedings of the 6th Conference on Robot Learning, 2023. 1769–1782
69 Wu F, Gu Z, Wu H, et al. Infer and adapt: bipedal locomotion reward learning from demonstrations via inverse reinforcement learning. In: Proceedings of 2024 IEEE International Conference on Robotics and Automation, 2024. 16243–16250
70 Belkhale S, Ding T, Xiao T, et al. Rt-h: action hierarchies using language. 2024. ArXiv:2403.01823
71 Kim M J, Pertsch K, Karamcheti S, et al. OpenVLA: an open-source vision-language-action model. In: Proceedings of the 8th Annual Conference on Robot Learning, 2024
72 Mees O, Ghosh D, Pertsch K, et al. Octo: an open-source generalist robot policy. In: Proceedings of the First Workshop on Vision-Language Models for Navigation and Manipulation at ICRA 2024, 2024
73 Intelligence P, Black K, Brown N, et al. $\pi_{0.5}$: a vision-language-action model with open-world generalization. 2025. ArXiv:2504.16054
74 Li Y, Wang X, Sun J, et al. Data-driven consensus control of fully distributed event-triggered multi-agent systems. Sci China Inf Sci, 2023, 66: 152202
75 Liu W, Sun J, Wang G, et al. Data-driven resilient predictive control under denial-of-service. IEEE Trans Automat Contr, 2023, 68: 4722–4737
76 Wang X, Sun J, Deng F, et al. Event-triggered consensus control of heterogeneous multi-agent systems: model- and data-based approaches. Sci China Inf Sci, 2023, 66: 192201
77 Wu J, Yang H, Gan T, et al. Chmatch: contrastive hierarchical matching and robust adaptive threshold boosted semi-supervised learning. In: Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023. 15762–15772
78 Wu J, Sun W, Gan T, et al. Neighbor-guided consistent and contrastive learning for semi-supervised action recognition. IEEE Trans Image Process, 2023, 32: 2215–2227
79 Zhao C, Xu J, Peng R, et al. Experience consistency distillation continual reinforcement learning for robotic manipulation tasks. In: Proceedings of 2024 IEEE International Conference on Robotics and Automation, 2024. 501–507
80 Zhao Z, Cheng S, Ding Y, et al. A survey of optimization-based task and motion planning: from classical to learning approaches. IEEE ASME Trans Mechatron, 2025, 30: 2799–2825
81 Lozano-Pérez T, Kaelbling L P. A constraint-based method for solving sequential manipulation planning problems. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014. 3684–3691
82 Dantam N T, Kingston Z K, Chaudhuri S, et al. An incremental constraint-based framework for task and motion planning. Int J Robotics Res, 2018, 37: 1134–1151
83 Yu F, Chen Y. Cyl-IRRT*: homotopy optimal 3D path planning for AUVs by biasing the sampling into a cylindrical informed subset. IEEE Trans Ind Electron, 2023, 70: 3985–3994
84 Ortiz-Haro J, Hönig W, Hartmann V N, et al. iDb-RRT: sampling-based kinodynamic motion planning with motion primitives and trajectory optimization. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2024. 10702–10709
85 Foehn P, Romero A, Scaramuzza D. Time-optimal planning for quadrotor waypoint flight. Sci Robot, 2021, 6: eabh1221
86 Mellinger D, Kumar V. Minimum snap trajectory generation and control for quadrotors. In: Proceedings of IEEE International Conference on Robotics and Automation, 2011. 2520–2525
87 Kulic D, Venture G, Bekris K, et al. MPCC++: model predictive contouring control for time-optimal flight with safety constraints. In: Proceedings of Robotics: Science and Systems Conference, 2024
88 Yang Z, Jia J, Liu Y, et al. TRACE: trajectory refinement with control error enables safe and accurate maneuvers. In: Proceedings of IEEE International Conference on Control & Automation, 2024. 154–161
89 Qin C, Michet M S, Chen J, et al. Time-optimal gate-traversing planner for autonomous drone racing. In: Proceedings of IEEE International Conference on Robotics and Automation, 2024. 8693–8699
90 Teissing K, Novosad M, Penicka R, et al. Real-time planning of minimum-time trajectories for agile UAV flight. IEEE Robot Autom Lett, 2024, 9: 10351–10358
91 Wang P, Tang J, Wang Lin H, et al. MINER-RRT*: a hierarchical and fast trajectory planning framework in 3D cluttered environments. IEEE Trans Automat Sci Eng, 2025, 22: 10973–10985
92 Fork T, Borrelli F. Euclidean and non-Euclidean trajectory optimization approaches for quadrotor racing. 2023. ArXiv:2309.07262

93 Ghotavadekar A, Nekovář F, Saska M, et al. Variable time-step MPC for agile multi-rotor UAV interception of dynamic targets. IEEE Robot Autom Lett, 2025, 10: 1249–1256
94 Zhou Z, Wang G, Sun J, et al. Efficient and robust time-optimal trajectory planning and control for agile quadrotor flight. IEEE Robot Autom Lett, 2023, 8: 7913–7920
95 Yu H, Hu C, Wang J, et al. Aggressive collision-inclusive motion planning. IEEE ASME Trans Mechatron, 2025, 30: 1412–1423
96 Wang G K, Liu Y H, Liu H P. One model, two skills: active vision and action learning model for robotic manipulation. Sci China Inf Sci, 2025, 68: 162202
97 Xiao W, Feng Z, Zhou Z, et al. Time-optimal flight in cluttered environments via safe reinforcement learning. Un Sys, 2026, 14: 391–400
98 Gao H, Chen S, Jiang B, et al. Rad: training an end-to-end driving policy via large-scale 3DGS-based reinforcement learning. 2025. ArXiv:2502.13144
99 Geng Y, An B, Geng H, et al. RLAfford: end-to-end affordance learning for robotic manipulation. In: Proceedings of IEEE International Conference on Robotics and Automation, 2023. 5880–5886
100 Bohlinger N, Czechmanowski G, Krupka M P, et al. One policy to run them all: an end-to-end learning approach to multi-embodiment locomotion. In: Proceedings of Annual Conference on Robot Learning, 2024
101 He T, Zhang C, Xiao W, et al. Agile but safe: learning collision-free high-speed legged locomotion. 2024. ArXiv:2401.17583
102 Hong C, Huang K, Liu H. Learning a distributed hierarchical locomotion controller for embodied cooperation. In: Proceedings of Annual Conference on Robot Learning, 2024
103 Jia B, Manocha D. Sim-to-real robotic sketching using behavior cloning and reinforcement learning. In: Proceedings of 2024 IEEE International Conference on Robotics and Automation, 2024. 18272-18278
104 Oh H, Matsubara T. Leveraging demonstrator-perceived precision for safe interactive imitation learning of clearance-limited tasks. IEEE Robot Autom Lett, 2024, 9: 3387–3394
105 Shi L X, Sharma A, Zhao T Z, et al. Waypoint-based imitation learning for robotic manipulation. In: Proceedings of the 7th Annual Conference on Robot Learning, 2023
106 Zakka K, Zeng A, Florence P, et al. Xirl: cross-embodiment inverse reinforcement learning. In: Proceedings of Conference on Robot Learning, 2022. 537–546
107 Wang Y, Liu Q, Jiang Z, et al. Rad: retrieval-augmented decision-making of meta-actions with vision-language models in autonomous driving. 2025. ArXiv:2503.13861
108 Chen L, Zhang Y, Ren S, et al. Towards end-to-end embodied decision making via multi-modal large language model: explorations with GPT4-vision and beyond. 2023. ArXiv:2310.02071
109 Jiang K, Cai X, Cui Z, et al. KoMA: knowledge-driven multi-agent framework for autonomous driving with large language models. IEEE Trans Intell Veh, 2025, 10: 4655–4668
110 Xu Z, Zhang Y, Xie E, et al. DriveGPT4: interpretable end-to-end autonomous driving via large language model. IEEE Robot Autom Lett, 2024, 9: 8186–8193
111 Huang Y, Sansom J, Ma Z, et al. Drivlme: enhancing llm-based autonomous driving agents with embodied and social experiences. In: Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2024. 3153–3160
112 Dong Z, Zhu Y, Li Y, et al. Generalizing end-to-end autonomous driving in real-world environments using zero-shot LLMs. In: Proceedings of the 8th Annual Conference on Robot Learning, 2024
113 Shao H, Hu Y, Wang L, et al. Lmdrive: closed-loop end-to-end driving with large language models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024. 15120–15130
114 Bjorck J, Castañeda F, Cherniadev N, et al. Gr00t n1: an open foundation model for generalist humanoid robots. 2025. ArXiv:2503.14734
115 Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. In: Proceedings of Advances in Neural Information Processing Systems, 2022. 24824–24837
116 Xie Y, Kawaguchi K, Zhao Y, et al. Self-evaluation guided beam search for reasoning. In: Proceedings of Advances in Neural Information Processing Systems, 2023. 41618–41650
117 Lin B Y, Fu Y, Yang K, et al. Swiftsage: a generative agent with fast and slow thinking for complex interactive tasks. In: Proceedings of Advances in Neural Information Processing Systems, 2023. 23813–23825
118 Hu Y, Lin F, Zhang T, et al. Look before you leap: unveiling the power of GPT-4v in robotic vision-language planning. 2023. ArXiv:2311.17842
119 Song C H, Wu J, Washington C, et al. LLM-planner: few-shot grounded planning for embodied agents with large language models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023. 2998–3009
120 Vemprala S H, Bonatti R, Bucker A, et al. ChatGPT for robotics: design principles and model abilities. IEEE Access, 2024, 12: 55682–55696
121 Zhang H, Zhu C, Wang X, et al. Badrobot: jailbreaking LLM-based embodied AI in the physical world. 2024. ArXiv:2407.20242
122 Yin S, Pang X, Ding Y, et al. Safeagentbench: a benchmark for safe task planning of embodied LLM agents. 2024. ArXiv:2412.13178
123 Wang L, Chen X, Zhao J, et al. Scaling proprioceptive-visual learning with heterogeneous pre-trained transformers. In: Proceedings of Advances in Neural Information Processing Systems, 2024. 124420–124450
124 Li L L, Yang B, Liang M, et al. End-to-end contextual perception and prediction with interaction transformer. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2020. 5784–5791
125 Mei Y, Sun J, Peng Z, et al. RoG-SAM: a language-driven framework for instance-level robotic grasping detection. IEEE Trans Multimedia, 2025, 27: 3057–3068
126 Li J, Liu X, Li B, et al. Comamba: real-time cooperative perception unlocked with state space models. 2024. ArXiv:2409.10699
127 Xu G, Ge Y, Liu M, et al. What matters when repurposing diffusion models for general dense perception tasks? 2024. ArXiv:2403.06090
128 Zhong Y, Kang J, Wen J, et al. Generative diffusion-based contract design for efficient AI twin migration in vehicular embodied AI networks. IEEE Trans Mobile Comput, 2025, 24: 4573–4588
129 Tian T, Li H, Ai B, et al. Diffusion dynamics models with generative state estimation for cloth manipulation. 2025. ArXiv:2503.11999
130 Kwon M, Xie S M, Bullard K, et al. Reward design with language models. In: Proceedings of the Eleventh International Conference on Learning Representations, 2023
131 Lee Y, Shin S, Park W J, et al. LLM-based offline learning for embodied agents via consistency-guided reward ensemble. 2024. ArXiv:2411.17135
132 Yu W, Gileadi N, Fu C, et al. Language to rewards for robotic skill synthesis. 2023. ArXiv:2306.08647
133 Ma Y J, Liang W, Wang G, et al. Eureka: human-level reward design via coding large language models. 2023. ArXiv:2310.12931
134 Xie T, Zhao S, Wu C H, et al. Text2reward: reward shaping with language models for reinforcement learning. 2023. ArXiv:2309.11489
135 Zeng Y, Mu Y, Shao L. Learning reward for robot skills using large language models via self-alignment. 2024. ArXiv:2405.07162
136 Sun S, Liu R, Lyu J, et al. A large language model-driven reward design framework via dynamic feedback for reinforcement learning. 2024. ArXiv:2410.14660
137 Zhang W, Wang G, Sun J, et al. STORM: efficient stochastic transformer based world models for reinforcement learning. In: Proceedings of Advances in Neural Information Processing Systems, 2023. 27147–27166
138 Robine J, Höftmann M, Uelwer T, et al. Transformer-based world models are happy with 100k interactions. In: Proceedings of the Eleventh International Conference on Learning Representations, 2023
139 Micheli V, Alonso E, Fleuret F. Transformers are sample-efficient world models. In: Proceedings of the Eleventh International Conference on Learning Representations, 2023

140 Hu A, Russell L, Yeo H, et al. Gaia-1: a generative world model for autonomous driving. 2023. ArXiv:2309.17080
141 Jia F, Mao W, Liu Y, et al. Adriver-i: a general world model for autonomous driving. 2023. ArXiv:2311.13549
142 Zhang L, Xiong Y, Yang Z, et al. Copilot4d: learning unsupervised world models for autonomous driving via discrete diffusion. 2023. ArXiv:2311.01017
143 Zhang Y, Gong S, Xiong K, et al. Bevworld: a multimodal world model for autonomous driving via unified bev latent space. 2024. ArXiv:2407.05679
144 Mazzaglia P, Verbelen T, Dhoedt B, et al. GenRL: multimodal-foundation world models for generalization in embodied agents. In: Proceedings of Advances in Neural Information Processing Systems, 2024
145 Wang S, Fei Z, Cheng Q, et al. World modeling makes a better planner: dual preference optimization for embodied task planning. 2025. ArXiv:2503.10480
146 Fung P, Bachrach Y, Celikyilmaz A, et al. Embodied AI agents: modeling the world. 2025. ArXiv:2506.22355
147 Nottingham K, Ammanabrolu P, Suhr A, et al. Do embodied agents dream of pixelated sheep: embodied decision making using language guided world modelling. In: Proceedings of International Conference on Machine Learning, 2023. 26311–26325
148 Wang Y, Xian Z, Chen F, et al. Robogen: towards unleashing infinite data for automated robot learning via generative simulation. 2023. ArXiv:2311.01455
149 Wang L, Ling Y, Yuan Z, et al. Gensim: generating robotic simulation tasks via large language models. 2023. ArXiv:2310.01361
150 Yang Y, Sun F Y, Weihs L, et al. Holodeck: language guided generation of 3S embodied AI environments. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024. 16227–16237
151 Zala A, Cho J, Lin H, et al. Envgen: generating and adapting environments via LLMs for training embodied agents. 2024. ArXiv:2403.12014
152 Shridhar M, Thomason J, Gordon D, et al. Alfred: a benchmark for interpreting grounded instructions for everyday tasks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020
153 Deitke M, Han W, Herrasti A, et al. Robothor: an open simulation-to-real embodied AI platform. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020
154 Chattopadhyay P, Hoffman J, Mottaghi R, et al. Robustnav: towards benchmarking robustness in embodied navigation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021. 15691–15700
155 Srivastava S, Li C, Lingelbach M, et al. Behavior: benchmark for everyday household activities in virtual, interactive, and ecological environments. In: Proceedings of the 5th Conference on Robot Learning, 2022. 477–490
156 Gu J, Xiang F, Li X, et al. Maniskill2: a unified benchmark for generalizable manipulation skills. In: Proceedings of the Eleventh International Conference on Learning Representations, 2023
157 Mu Y, Zhang Q, Hu M, et al. EmbodiedGPT: vision-language pre-training via embodied chain of thought. In: Proceedings of Advances in Neural Information Processing Systems, 2023. 25081–25094
158 Sferrazza C, Huang D M, Lin X, et al. Humanoidbench: simulated humanoid benchmark for whole-body locomotion and manipulation. 2024. ArXiv:2403.10506
159 O'Neill A, Rehman A, Maddukuri A, et al. Open x-embodiment: robotic learning datasets and rt-x models: open x-embodiment collaboration. In: Proceedings of 2024 IEEE International Conference on Robotics and Automation, 2024. 6892–6903
160 Majumdar A, Ajay A, Zhang X, et al. Openeqa: embodied question answering in the era of foundation models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024. 16488–16498
161 Li M, Zhao S, Wang Q, et al. Embodied agent interface: benchmarking LLMs for embodied decision making. In: Proceedings of Advances in Neural Information Processing Systems, 2024. 100428–100534
162 Kim T, Min C, Kim B, et al. ReALFRED: an embodied instruction following benchmark in photo-realistic environments. In: Proceedings of Computer Vision – ECCV 2024, 2025. 346–364
163 Chow W, Mao J, Li B, et al. Physbench: benchmarking and enhancing vision-language models for physical world understanding. 2025. ArXiv:2501.16411
164 Xiang F, Qin Y, Mo K, et al. Sapien: a simulated part-based interactive environment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020
165 Brockman G, Cheung V, Pettersson L, et al. OpenAI gym. 2016. ArXiv:1606.01540
166 Todorov E, Erez T, Tassa Y. Mujoco: a physics engine for model-based control. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012. 5026–5033
167 Ramos S, Girgin S, Hussenot L, et al. Rlds: an ecosystem to generate, share and use datasets in reinforcement learning. 2021. ArXiv:2111.02767
168 Yu C, Yang X, Gao J, et al. Asynchronous multi-agent reinforcement learning for efficient real-time multi-robot cooperative exploration. In: Proceedings of the International Conference on Autonomous Agents and Multiagent Systems, 2023. 1107–1115
169 Bettini M, Shankar A, Prorok A. Heterogeneous multi-robot reinforcement learning. In: Proceedings of AAMAS, 2023. 1485–1494
170 Gao Y, Chen J, Chen X, et al. Asymmetric self-play-enabled intelligent heterogeneous multirobot catching system using deep multiagent reinforcement learning. IEEE Trans Robot, 2023, 39: 2603–2622
171 Singh I, Traum D, Thomason J. Twostep: multi-agent task planning using classical planners and large language models. 2024. ArXiv:2403.17246
172 Venkatesh V L, Min B C. Zerocap: zero-shot multi-robot context aware pattern formation via large language models. 2024. ArXiv:2404.02318
173 Liu K, Tang Z, Wang D, et al. Coherent: collaboration of heterogeneous multi-robot system with large language models. 2024. ArXiv:2409.15146
174 Zhang B, Mao H, Ruan J, et al. Controlling large language model-based agents for large-scale decision-making: an actor-critic approach. In: Proceedings of ICLR 2024 Workshop on Large Language Model (LLM) Agents, 2024
175 Liu J, Chen M. Fagel: fabric llms agent empowered embodied intelligence evolution with autonomous human-machine collaboration. 2024. ArXiv:2412.20297
176 Qian C, Xie Z, Wang Y, et al. Scaling large language model-based multi-agent collaboration. In: Proceedings of the Thirteenth International Conference on Learning Representations, 2025
177 Zhang H, Du W, Shan J, et al. Building cooperative embodied agents modularly with large language models. In: Proceedings of the Twelfth International Conference on Learning Representations, 2024
178 Wan Y, Wu Y, Wang Y, et al. Infer human's intentions before following natural language instructions. AAAI, 2025, 39: 25309–25317
179 Yuan P, Ma A, Yao Y, et al. Remac: self-reflective and self-evolving multi-agent collaboration for long-horizon robot manipulation. 2025. ArXiv:2503.22122
180 Liu J, Xu C, Hang P, et al. Language-driven policy distillation for cooperative driving in multi-agent reinforcement learning. IEEE Robot Autom Lett, 2025, 10: 4292–4299
181 Silver D. Cooperative pathfinding. AIIDE, 2021, 1: 117–122
182 Augugliaro F, Schoellig A P, D'Andrea R. Generation of collision-free trajectories for a quadrocopter fleet: a sequential convex programming approach. In: Proceedings of 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012. 1917–1922
183 Hou J, Zhou X, Gan Z, et al. Enhanced decentralized autonomous aerial robot teams with group planning. IEEE Robot Autom Lett, 2022, 7: 9240–9247
184 Tordesillas J, How J P. MADER: trajectory planner in multiagent and dynamic environments. IEEE Trans Robot, 2022, 38: 463–476

185 Quan L, Yin L, Xu C, et al. Distributed swarm trajectory optimization for formation flight in dense environments. In: Proceedings of 2022 International Conference on Robotics and Automation, 2022. 4979–4985
186 Xiao Y, Tan W, Amato C. Asynchronous actor-critic for multi-agent reinforcement learning. In: Proceedings of Advances in Neural Information Processing Systems, 2022. 4385–4400
187 Xiao Y, Tan W, Hoffman J, et al. Asynchronous multi-agent deep reinforcement learning under partial observability. Int J Robotics Res, 2025, 44: 1257–1286
188 Cai Y, He X, Guo H, et al. Transformer-based multi-agent reinforcement learning for generalization of heterogeneous multi-robot cooperation. In: Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2024. 13695–13702
189 Zhong Y, Kuba J G, Feng X, et al. Heterogeneous-agent reinforcement learning. J Mach Learn Res, 2024, 25: 1–67
190 Yuan L, Jiang T, Li L H, et al. Robust cooperative multi-agent reinforcement learning via multi-view message certification. Sci China Inf Sci, 2024, 67: 142102
191 Yuan L, Zhang Z, Xue K, et al. Robust multi-agent coordination via evolutionary generation of auxiliary adversarial attackers. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2023. 11753–11762
192 Yuan L, Li L, Zhang Z, et al. Multiagent continual coordination via progressive task contextualization. IEEE Trans Neural Netw Learn Syst, 2025, 36: 6326–6340
193 Kannan S S, Venkatesh V L N, Min B C. Smart-llm: smart multi-agent robot task planning using large language models. In: Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2024. 12140–12147
194 Chen W, Koenig S, Dilkina B. Why solving multi-agent path finding with large language model has not succeeded yet. 2024. ArXiv:2401.03630
195 Wan H, Chen Y, Wei Z, et al. Embodiedagent: a scalable hierarchical approach to overcome practical challenge in multi-robot control. 2025. ArXiv:2504.10030
196 Wu Y, Tao Y, Li P, et al. Hierarchical LLMs in-the-loop optimization for real-time multi-robot target tracking under unknown hazards. 2024. ArXiv:2409.12274
197 Chen J, Yu C, Zhou X, et al. EMOS: embodiment-aware heterogeneous multi-robot operating system with LLM agents. In: Proceedings of the Thirteenth International Conference on Learning Representations, 2025
198 Wang Y, Xiao R, Kasahara J Y L, et al. Dart-llm: dependency-aware multi-robot task decomposition and execution using large language models. 2024. ArXiv:2411.09022
199 Obata K, Aoki T, Horii T, et al. LiP-LLM: integrating linear programming and dependency graph with large language models for multi-robot task planning. IEEE Robot Autom Lett, 2025, 10: 1122–1129
200 Zhang Y, Yang S, Bai C, et al. Towards efficient LLM grounding for embodied multi-agent collaboration. 2024. ArXiv:2405.14314
201 Wang W, Obi I, Min B C. Multi-agent LLM actor-critic framework for social robot navigation. 2025. ArXiv:2503.09758
202 Bo X, Zhang Z, Dai Q, et al. Reflective multi-agent collaboration based on large language models. In: Proceedings of Advances in Neural Information Processing Systems, 2024. 138595–138631
203 Guo X, Huang K, Liu J, et al. Embodied LLM agents learn to cooperate in organized teams. In: Proceedings of Language Gamification - NeurIPS 2024 Workshop, 2024
204 Tan S, Xiang W, Liu H, et al. Multi-agent embodied question answering in interactive environments. In: Proceedings of Computer Vision – ECCV 2020, 2020. 663–678
205 Wang J, He G, Kantaros Y. Safe task planning for language-instructed multi-robot systems using conformal prediction. 2024. ArXiv:2402.15368
206 Yu W, Peng J, Ying Y, et al. Mhrc: closed-loop decentralized multi-heterogeneous robot collaboration with large language models. 2024. ArXiv:2409.16030
207 Liu Z, Zhang Y, Li P, et al. Dynamic LLM-agent network: an LLM-agent collaboration framework with agent team optimization. 2023. ArXiv:2310.02170
208 Chen H, Ji W, Xu L, et al. Multi-agent consensus seeking via large language models. 2023. ArXiv:2310.20151
209 Ajoudani A, Zanchettin A M, Ivaldi S, et al. Progress and prospects of the human-robot collaboration. Auton Robot, 2018, 42: 957–975
210 Vicentini F. Collaborative robotics: a survey. J Mech Des, 2021, 143: 040802
211 van Berkel N, Skov M B, Kjeldskov J. Human-AI interaction: intermittent, continuous, and proactive. Interactions, 2021, 28: 67–71
212 Onnasch L, Roesler E. A taxonomy to structure and analyze human-robot interaction. Int J Soc Robotics, 2021, 13: 833–849
213 Sun N, Shi C, Dong Y. Leveraging LLM-based multi-agent collaboration to enhance embodied reasoning in human-populated environments. In: Proceedings of Submitted to Tsinghua University Course: Advanced Machine Learning & Machine Learning, 2024
214 Feng X, Chen Z Y, Qin Y, et al. Large language model-based human-agent collaboration for complex task solving. 2024. ArXiv:2402.12914
215 Guan C, Zhang L, Fan C, et al. Efficient human-AI coordination via preparatory language-based convention. 2023. ArXiv:2311.00416
216 Asuzu K, Singh H, Idrissi M. Human-robot interaction through joint robot planning with large language models. Intel Serv Robotics, 2025, 18: 261–277
217 Liu J, Zhou P, Du Y, et al. Capo: cooperative plan optimization for efficient embodied multi-agent cooperation. 2024. ArXiv:2411.04679
218 Gao J, Wang Z, Xiao Z, et al. Coohoi: learning cooperative human-object interaction with manipulated object dynamics. In: Proceedings of Advances in Neural Information Processing Systems, 2024. 79741–79763
219 Francesco T, Edoardo Z, Gianni F, et al. Collaborative instance object navigation: leveraging uncertainty-awareness to minimize human-agent dialogues. 2025. ArXiv:2412.01250
220 Krupnik O, Mordatch I, Tamar A. Multi-agent reinforcement learning with multi-step generative models. In: Proceedings of the Conference on Robot Learning, 2020. 776–790
221 Shi Z, Liu M, Zhang S, et al. Gawm: global-aware world model for multi-agent reinforcement learning. 2025. ArXiv:2501.10116
222 Venugopal A, Milani S, Fang F, et al. Mabl: bi-level latent-variable world model for sample-efficient multi-agent reinforcement learning. In: Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems, 2024. 1865–1873
223 Barde P, Foerster J, Nowrouzezahrai D, et al. A model-based solution to the offline multi-agent reinforcement learning coordination problem. In: Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems, 2024. 141–150
224 Liu Z, Yang X, Sun S, et al. Grounded answers for multi-agent decision-making problem through generative world model. In: Proceedings of Advances in Neural Information Processing Systems, 2024. 46622–46652
225 Zhang Z, Liniger A, Dai D, et al. Trafficbots: towards world models for autonomous driving simulation and motion prediction. In: Proceedings of 2023 IEEE International Conference on Robotics and Automation, 2023. 1522–1529
226 Zhang H, Wang Z, Lyu Q, et al. COMBO: compositional world models for embodied multi-agent cooperation. In: Proceedings of the Thirteenth International Conference on Learning Representations, 2025
227 Stern R, Sturtevant N, Felner A, et al. Multi-agent pathfinding: definitions, variants, and benchmarks. In: Proceedings of the International Symposium on Combinatorial Search, 2019. 151–158
228 Jain U, Weihs L, Kolve E, et al. A cordial sync: going beyond marginal policies for multi-agent embodied tasks. In: Proceedings of Computer Vision – ECCV 2020, 2020. 471–490
229 Zhou M, Luo J, Villella J, et al. Smarts: an open-source scalable multi-agent RL training school for autonomous driving. In: Proceedings of the 2020 Conference on Robot Learning, 2021. 264–285
230 Petrenko A, Wijmans E, Shacklett B, et al. Megaverse: simulating embodied agents at one million experiences per second. In: Proceedings of the 38th International Conference on Machine Learning, 2021. 8556–8566
231 Li Y, Ma D, An Z, et al. V2X-Sim: multi-agent collaborative perception dataset and benchmark for autonomous driving. IEEE Robot

Autom Lett, 2022, 7: 10914–10921
232 Chen Y, Wu T, Wang S, et al. Towards human-level bimanual dexterous manipulation with reinforcement learning. In: Proceedings of Advances in Neural Information Processing Systems, 2022. 5150–5163
233 Schaefer S, Palmieri L, Heuer L, et al. A benchmark for multi-robot planning in realistic, complex and cluttered environments. In: Proceedings of 2023 IEEE International Conference on Robotics and Automation, 2023. 9231–9237
234 Gong R, Gao X, Gao Q, et al. LEMMA: learning language-conditioned multi-robot manipulation. IEEE Robot Autom Lett, 2023, 8: 6835–6842
235 Du W, Lyu Q, Shan J, et al. Constrained human-AI cooperation: an inclusive embodied social intelligence challenge. In: Proceedings of Advances in Neural Information Processing Systems, 2024
236 Mandi Z, Jain S, Song S. Roco: dialectic multi-robot collaboration with large language models. In: Proceedings of 2024 IEEE International Conference on Robotics and Automation, 2024. 286–299
237 Nayak S, Orozco A M, Have M T, et al. Long-horizon planning for multi-agent robots in partially observable environments. In: Proceedings of Advances in Neural Information Processing Systems, 2024. 67929–67967
238 Heuer L, Palmieri L, Mannucci A, et al. Benchmarking multi-robot coordination in realistic, unstructured human-shared environments. In: Proceedings of 2024 IEEE International Conference on Robotics and Automation, 2024. 14541–14547
239 Zhang H, Chen J, Cheng J, et al. Nowhere to go: benchmarking multirobot collaboration in target trapping environment. IEEE Trans Ind Electron, 2024, 71: 14430–14439
240 Chandra R, Sprague Z, Biswas J. SOCIALGYM 2.0: simulator for multi-robot learning and navigation in shared human spaces. AAAI, 2024, 38: 23778–23780
241 Blumenkamp J, Shankar A, Bettini M, et al. The cambridge robomaster: an agile multi-robot research platform. 2024. ArXiv:2405.02198
242 Feng D, Qi Y, Zhong S, et al. S3E: a multi-robot multimodal dataset for collaborative SLAM. IEEE Robot Autom Lett, 2024, 9: 11401–11408
243 Chang M, Chhablani G, Clegg A, et al. PARTNR: a benchmark for planning and reasoning in embodied multi-agent tasks. In: Proceedings of the Thirteenth International Conference on Learning Representations, 2025
244 Yang R, Chen H, Zhang J, et al. Embodiedbench: comprehensive benchmarking multi-modal large language models for vision-driven embodied agents. 2025. ArXiv:2502.09560
245 Hartmann V N, Heinle T, Coros S. A benchmark for optimal multi-modal multi-robot multi-goal path planning with given robot assignment. 2025. ArXiv:2503.03509
246 Cao M, Nguyen T M, Yuan S, et al. Cooperative aerial robot inspection challenge: a benchmark for heterogeneous multi-uav planning and lessons learned. 2025. ArXiv:2501.06566
247 Makoviychuk V, Wawrzyniak L, Guo Y, et al. Isaac Gym: high performance GPU based physics simulation for robot learning. In: Proceedings of Advances in Neural Information Processing Systems, 2021
248 Calli B, Singh A, Walsman A, et al. The ycb object and model set: towards common benchmarks for manipulation research. In: Proceedings of 2015 International Conference on Advanced Robotics (ICAR), 2015. 510–517
249 Macenski S, Foote T, Gerkey B, et al. Robot operating system 2: design, architecture, and uses in the wild. Sci Robot, 2022, 7: eabm6074
250 Koenig N, Howard A. Design and use paradigms for Gazebo, an open-source multi-robot simulator. In: Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004. 2149–2154
251 Terry J, Black B, Grammel N, et al. Pettingzoo: gym for multi-agent reinforcement learning. In: Proceedings of Advances in Neural Information Processing Systems, 2021. 15032–15043
252 Bettini M, Kortvelesy R, Blumenkamp J, et al. Vmas: a vectorized multi-agent simulator for collective robot learning. In: Proceedings of Distributed Autonomous Robotic Systems, 2024. 42–56
253 Sunehag P, Lever G, Gruslys A, et al. Value-decomposition networks for cooperative multi-agent learning. 2017. ArXiv:1706.05296
254 Foerster J N, Farquhar G, Afouras T, et al. Counterfactual multi-agent policy gradients. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 2018. 2974–2982
255 Chu T, Chinchali S, Katti S. Multi-agent reinforcement learning for networked system control. In: Proceedings of International Conference on Learning Representations, 2020
256 Strachan J W A, Albergo D, Borghini G, et al. Testing theory of mind in large language models and humans. Nat Hum Behav, 2024, 8: 1285–1295
257 Amato C. An introduction to centralized training for decentralized execution in cooperative multi-agent reinforcement learning. 2024. ArXiv:2409.03052
258 Yu Y. Towards sample efficient reinforcement learning. In: Proceedings of International Joint Conferences on Artificial Intelligence, 2018. 5739–5743
259 Cao Y, Zhao H, Cheng Y, et al. Survey on large language model-enhanced reinforcement learning: concept, taxonomy, and methods. IEEE Trans Neural Netw Learn Syst, 2025, 36: 9737–9757
260 Luo F-M, Xu T, Lai H, et al. A survey on model-based reinforcement learning. Sci China Inf Sci, 2024, 67: 121101
261 Jiang T, Yuan L, Li L, et al. Multi-agent domain calibration with a handful of offline data. In: Proceedings of the Thirty-eighth Annual Conference on Neural Information Processing Systems, 2025
262 Huang D, Yan C, Li Q, et al. From large language models to large multimodal models: a literature review. Appl Sci, 2024, 14: 5068
263 Reed S, Zolna K, Parisotto E, et al. A generalist agent. 2022. ArXiv:2205.06175
264 Wen M, Kuba J, Lin R, et al. Multi-agent reinforcement learning is a sequence modeling problem. In: Proceedings of Advances in Neural Information Processing Systems, 2022. 16509–16521
265 Zhou Z H. Open-environment machine learning. Natl Sci Rev, 2022, 9: nwac123
266 Farias K, Moraes P, Nunes I, et al. Robocup Rescue 2025 team description paper urubots. 2025. ArXiv:2504.09778
267 Samvelyan M, Rashid T, De Witt C S, et al. The StarCraft multi-agent challenge. 2019. ArXiv:1902.04043
268 Liu Y, Cao X, Chen T, et al. From screens to scenes: a survey of embodied AI in healthcare. Inf Fusion, 2025, 119: 103033
269 Ma Y, Ye W, Cui C, et al. Position: prospective of autonomous driving-multimodal LLMs, world models, embodied intelligence, AI alignment, and mamba. In: Proceedings of 2025 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, 2025. 920–936
270 Berner C, Brockman G, Chan B, et al. Dota 2 with large scale deep reinforcement learning. 2019. ArXiv:1912.06680
271 Vinyals O, Babuschkin I, Czarnecki W M, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature, 2019, 575: 350–354
272 Duan J, Zhuang L, Zhang Q, et al. Multimodal perception-fusion-control and human-robot collaboration in manufacturing: a review. Int J Adv Manuf Technol, 2024, 132: 1071–1093
273 Ndousse K K, Eck D, Levine S, et al. Emergent social learning via multi-agent reinforcement learning. In: Proceedings of International Conference on Machine Learning, 2021. 7991–8004
274 Ruane E, Birhane A, Ventresque A. Conversational AI: social and ethical considerations. AICS, 2019, 2563: 104–115