
MULTI-AGENT EMBODIED AI: ADVANCES AND FUTURE DIRECTIONS

A PREPRINT

Zhaohan Feng¹, Ruiqi Xue², Lei Yuan², Yang Yu², Ning Ding³, Meiqin Liu^{3,4}, Bingzhao Gao⁵, Jian Sun¹, Xinhua Zheng⁶, and 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 210000, China

³ State 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

zhfeng@bit.edu.cn, xuerq@lamda.nju.edu.cn, yuanl@lamda.nju.edu.cn,
yuy@nju.edu.cn, ding.ning@xjtu.edu.cn, liumeiqin@zju.edu.cn,
gaobz@jlu.edu.cn, sunjian@bit.edu.cn, xinhuzheng@hkust-gz.edu.cn, gangwang@bit.edu.cn

June 24, 2025

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) mature, 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 in the field, this paper 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.

*Corresponding Author

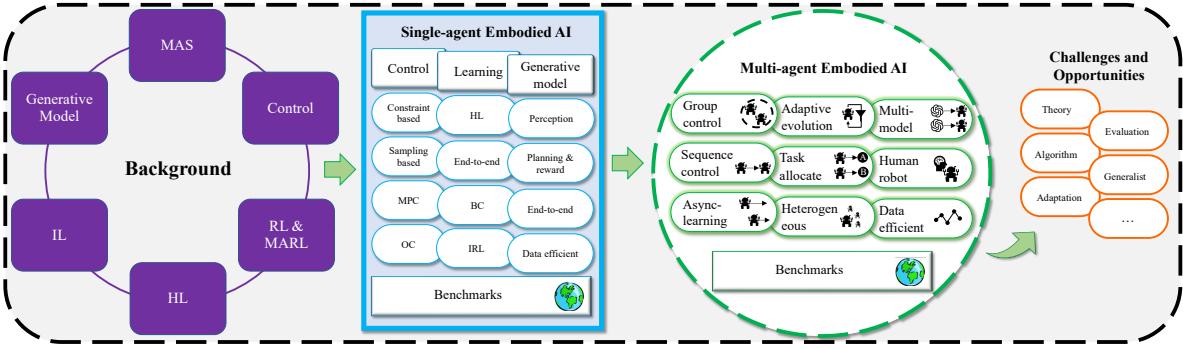


Figure 1: An overview of this survey.

1 Introduction

Embodied artificial intelligence (Embodied AI) [144, 40] is an interdisciplinary research area at the intersection of artificial intelligence (AI), robotics, and cognitive science, aimed at equipping robots with the capabilities to perceive, plan, decide, and act, 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 [184]. 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 learning, thus laying the foundation for embodied AI [18]. 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 large pre-trained models such as large language models (LLMs) and vision-language models (VLMs). These models have enhanced intelligent systems’ capabilities in perception, decision-making, and task planning, achieving outstanding performance in visual understanding, natural language processing, and multi-modal integration tasks [201]. Leveraging these large models, embodied AI has achieved breakthroughs in image recognition, speech interaction, and robotic control, improving robots’ ability to understand and adapt to complex dynamic environments [111].

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 [154]. 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 [132]. 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 [17]. 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 [243]. As an alternative, RL, especially deep RL, integrates deep neural networks to optimize decision-making strategies through interaction-driven feedback from the environment [43]. 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 [221] and VLMs [117], has significantly advanced embodied AI, enhancing agents’ capabilities in visual understanding, multi-modal 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 [37].

Multi-agent settings fundamentally differ from single-agent scenarios, as agents must simultaneously optimize their individual policies and manage complex interactions among multiple entities. Specifically, multi-agent interactions introduce challenges such as exponential growth in problem complexity due to enlarged joint action spaces and extended planning horizons, partial observability stemming from decentralized information among agents, non-stationarity arising from concurrent agent learning processes, and difficulties associated with accurately assigning credit to individual contributions [251, 238, 2]. Despite substantial advancements in single-agent embodied AI, research on embodied AI

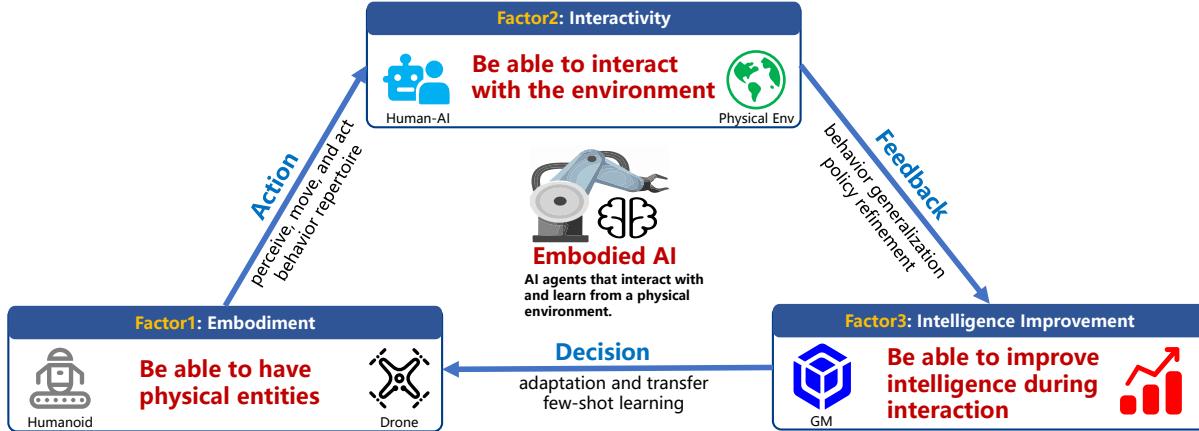


Figure 2: An illustration of embodied AI.

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 [144, 111], multi-agent reinforcement learning (MARL) [251, 2], and multi-agent cooperation [238, 138], comprehensive surveys explicitly focusing on embodied multi-agent AI remain limited. For example, Literature [178] systematically summarizes recent progress in embodied MARL, covering topics such as social learning, emergent communication, Sim2Real transfer, hierarchical methods, and safety considerations. Literature [209] further extends 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 paper 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 summarize the primary contributions of this review, present insightful perspectives on the future development of multi-agent embodied AI, and aim to stimulate further research and innovation in this promising and 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 [111]. The concept of embodied AI trace 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 [184]. 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 [18]. 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.



Figure 3: Three common settings of MAS.

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. Then 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 [15], which enables embodied agents to adapt their behavior dynamically based on feedback. The paradigm of embodied AI mainly include 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.

Recently, the advancements in generative models such as LLMs [258] 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 [117]. Importantly, embodied AI is not merely the integration of powerful AI models with robotic platforms. It rather 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 [37]. 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. Core attributes of MAS include autonomy, decentralization, inter-agent communication, local information access, and dynamic adaptability. These features collectively empower MAS to tackle a

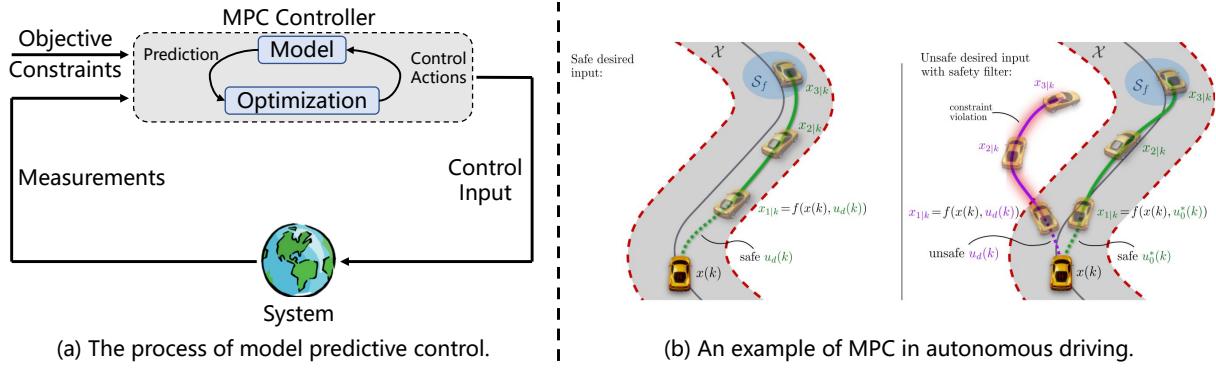


Figure 4: An illustration of MPC.

broad spectrum of complex, high-dimensional tasks that 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 [2], 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 [63].

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 [153]. 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:

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

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:

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

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 (UAV).

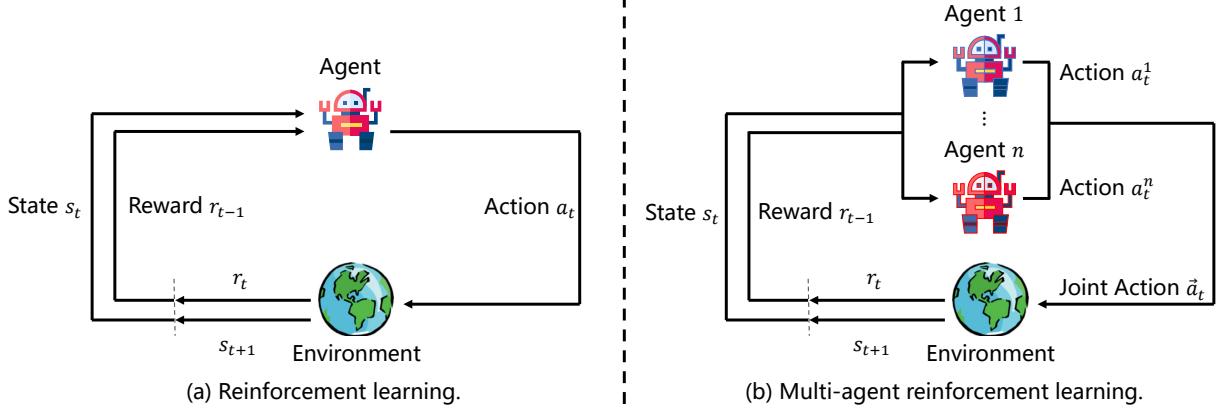


Figure 5: An illustration of RL and MARL.

2.4 Reinforcement Learning

Reinforcement Learning (RL) is a fundamental branch of machine learning that focuses on enabling agents to make sequential decisions by interacting with dynamic environments [176]. 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 [94, 169], robotic control [165], industrial process optimization [133], personalized healthcare [228], autonomous driving [89, 103], and even alignment and instruction-following in LLMs [140, 66].

At its core, an RL problem is commonly formalized as a Markov decision process (MDP), depicted in Figure 5(a). An MDP is defined by 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 a 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

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

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 a via $P(s'|s, a)$, and all agents receive a shared reward $R(s, a)$. Observations are generated through the function $O(s', a, o)$, providing each agent with private, potentially incomplete information. The central challenge lies in learning a set of decentralized policies $[\pi_1, \dots, \pi_n]$, 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 [21, 264, 85]. Among these, two principal categories stand out: value-based methods and policy-based methods. Value-based methods, such as the deep Q-network (DQN) [128], 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 [104] and PPO [158], 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 [149], 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 [114] and MAPPO [229], adopt the CTDE framework. These methods employ a

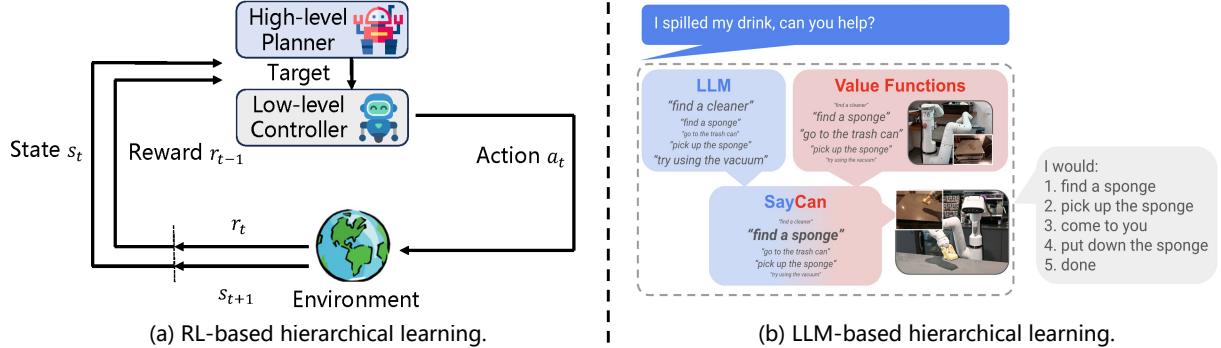


Figure 6: An illustration of RL-based (high-level) hierarchical learning and LLM-based (high-level) hierarchical learning [78].

shared global critic during training to provide stabilized learning signals and support effective credit assignment among agents, thereby enhancing coordination in cooperative tasks.

2.5 Hierarchical Learning

Hierarchical learning is a learning 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 (cf. Figure 6). 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 [142].

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 classic controllers such as MPC [168, 47, 42], or through end-to-end RL [224, 98]. 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 [224, 98]. 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 [78].

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) [243].

Among these, BC is the most straightforward and widely adopted technique. It casts the imitation task as a supervised learning problem, aiming to directly map observed states to corresponding expert actions. As illustrated in Figure 7(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:

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

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 error accumulation 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 7(b), enables

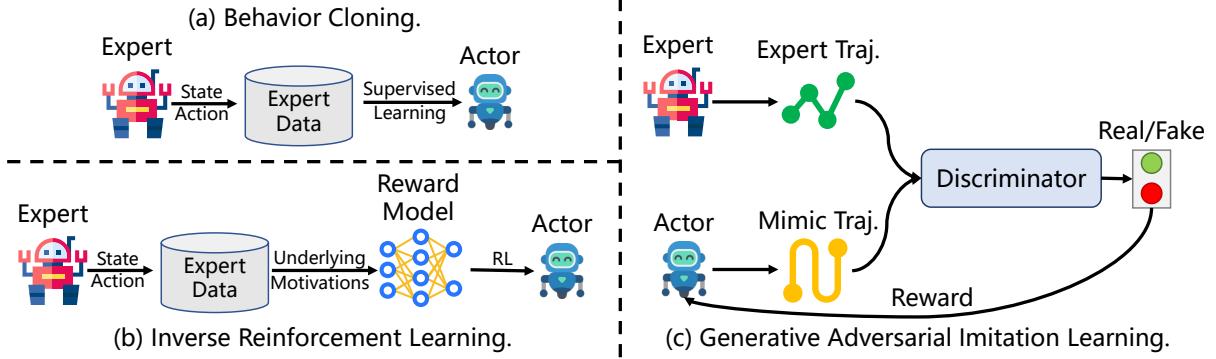


Figure 7: An illustration of different IL methods.

improved generalization and policy interpretability. Formally, a representative formulation is the maximum entropy IRL objective

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

where $\tau = (s_0, a_0, \dots, s_T, a_T)$, $R(\tau) = \sum_{t=0}^T R(s_t, a_t)$, 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 [69] introduces an adversarial training framework that eliminates the need for explicit reward specification. Inspired by generative adversarial networks (GANs) [59], 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 7(c), the optimization objective is

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

where \mathcal{D}_{π} 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 multi-modal 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 [186], diffusion models [70] 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 8(a)). This design facilitates efficient parallel computation and allows the model to capture global contextual dependencies. The core attention mechanism is mathematically defined as:

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

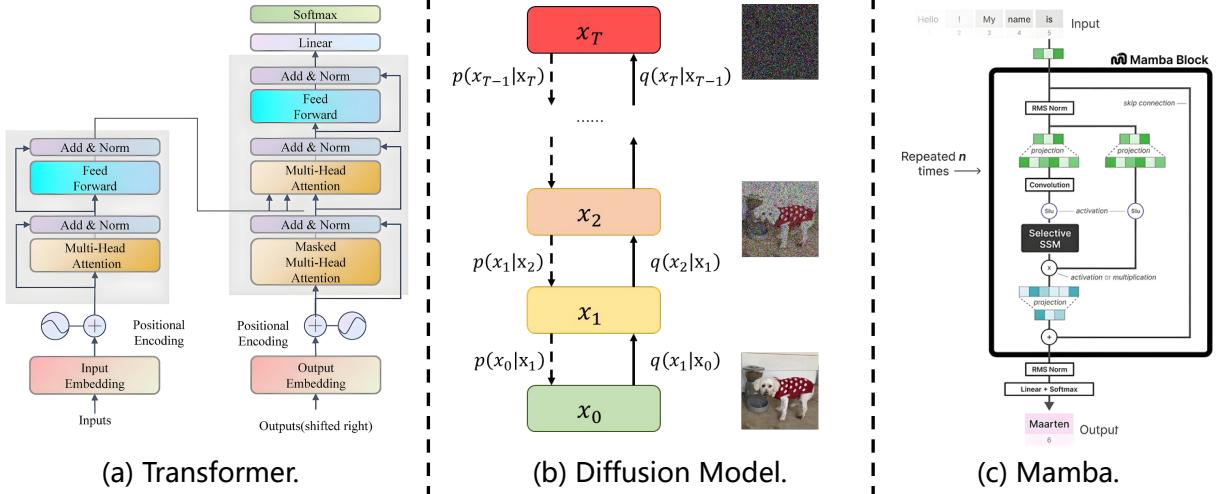


Figure 8: An illustration of different generative model structures.

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 8(b)). The forward diffusion process forms a Markov chain in which Gaussian noise is added at each timestep:

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

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:

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

where μ_θ and Σ_θ are parameterized by neural networks, typically based on U-Net architectures [152]. 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 8(c)). Mamba represents sequential data using the following differential equations:

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

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.

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

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				
RRT [95]	1998		✓				✓		✓	✓
MINCO [206]	2022		✓	✓	✓		✓	✓	✓	✓
High-MPC [168]	2022		✓	✓	✓		✓	✓	✓	✓
SteveEye [260]	2023	✓	✓	✓	✓	✓	✓	✓	✓	✓
SayCan [78]	2023	✓	✓	✓	✓	✓	✓	✓	✓	✓
VisionRacing [52]	2023	✓	✓	✓	✓	✓	✓	✓	✓	✓
GNFactor [244]	2023	✓	✓	✓	✓	✓	✓	✓	✓	✓
PaLM-E [38]	2023	✓	✓	✓	✓	✓	✓	✓	✓	✓
InnerMonologue [76]	2023	✓	✓		✓	✓	✓	✓	✓	✓
Infer & Adapt [210]	2024			✓		✓	✓	✓		✓
RT-H [7]	2024	✓	✓	✓	✓	✓	✓	✓	✓	✓
OpenVLA [87]	2024	✓	✓	✓	✓	✓	✓	✓	✓	✓
Octo [124]	2024	✓	✓	✓	✓	✓	✓	✓	✓	✓
$\pi_{0.5}$ [79]	2025	✓	✓	✓	✓	✓	✓	✓	✓	✓

categorized into classic methods and modern learning-based techniques. Early efforts primarily relied on knowledge-driven, task-specific planning and control strategies [102, 110, 202]. 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 [212, 211, 257]. 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 Classic 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 classic planning methods has been developed, which can be broadly grouped into constraint-based, sampling-based, and optimization-based approaches [259]. 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 [115, 34]. 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 [95, 231, 139].

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 [126], MPC [92] and OC [48]. 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 [226, 146, 179, 198, 50]. In contrast, MPC and OC approaches are better suited for highly constrained scenarios requiring time-optimal maneuvers, such as collision avoidance during UAV operations [48, 57, 265, 232]. For example, as illustrated in Figure 9, the work [232] 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

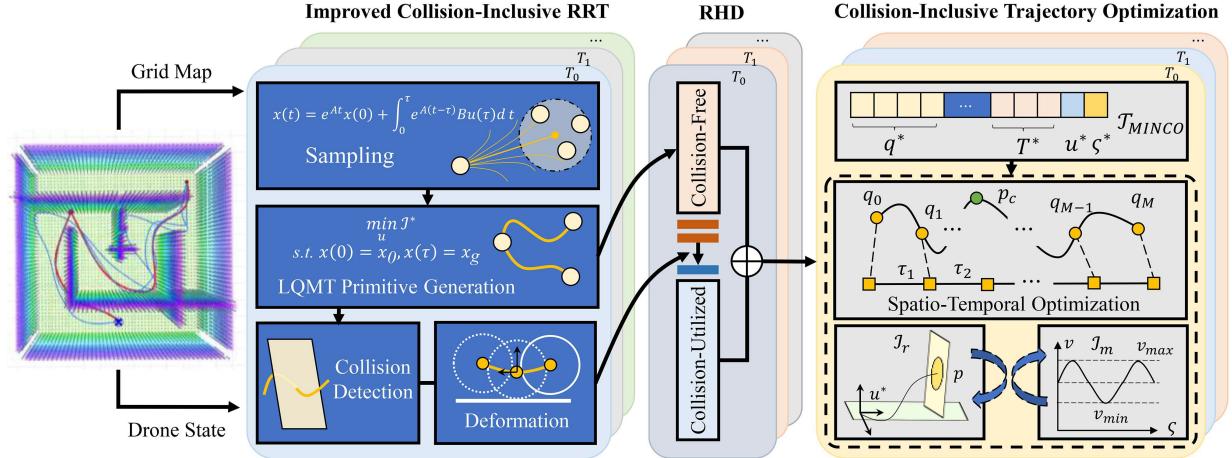


Figure 9: An illustration of control-based motion planning [232].

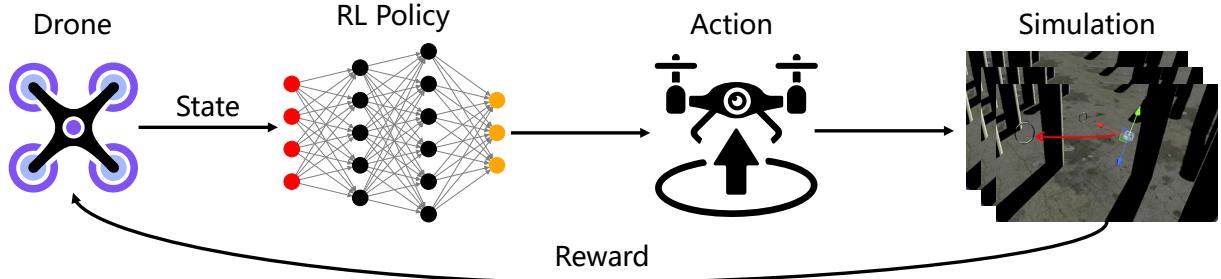


Figure 10: An illustration of end-to-end RL methods, where the learned policy π directly interacts with the environment [215].

(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

Classic 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 [194]. 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 classic 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

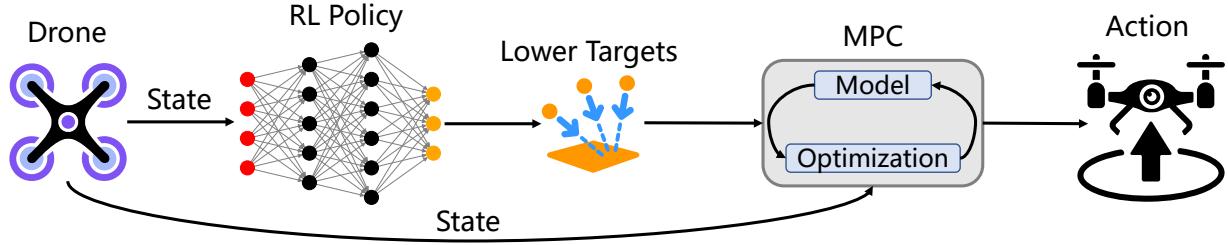


Figure 11: An illustration of hierarchical learning using the example of embodied drone control.



Figure 12: An illustration of the writing and painting skills learned from demonstrations [81].

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 (cf. Figure 10). 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 multi-modal inputs, including UAVs [215, 52], autonomous vehicles [53], manipulators [56] and legged robots [14, 67].

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 classic 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 [168, 47, 71]. 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 classic methods alone (Figure 11), as the gate’s position evolves overtime, 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 timing to generate efficient trajectories [168]. 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 [47]. Additionally, carefully structured hierarchical designs can naturally extend this framework to support multi-agent collaborative tasks in complex embodied scenarios [71].

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, [81] successfully transfers a painting policy from a simulator to a real-world robotic arm using BC, demonstrating high-level painting and calligraphy skills as shown in Figure 12. 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 [136], constrained IL [161], and IRL [241].

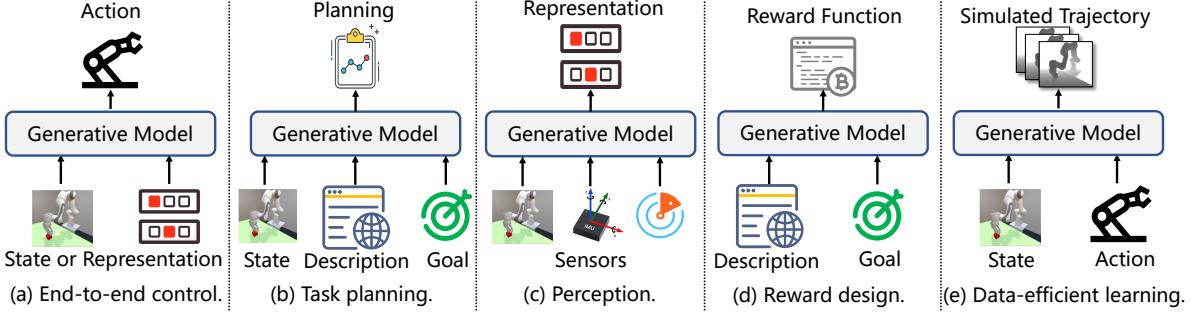


Figure 13: An illustration of different roles generative models played in embodied control.

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 [53, 203]. 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, multi-modal 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 13(a). Early approaches typically involve formatting inputs such as states and goals in natural language and feed them into pre-trained generative models, enabling action inference without additional training [29, 83]. 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 [222, 77, 38, 87, 124, 79], or integrating them with additional networks [36, 160, 11] 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 [207] and reflection capabilities [219, 105, 74, 76] 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 [78, 167, 187], as shown in Figure 13(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 [250, 227].

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 [196] 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 13(c). Similarly, numerous methods leverage Transformer, diffusion Model, and Mamba architectures for perceptual fusion [99, 125, 97, 220, 261, 181]. Additionally, leveraging pre-trained VLMs for semantic understanding of multimodal perception data provides another effective approach to support embodied agents in perception tasks [260].

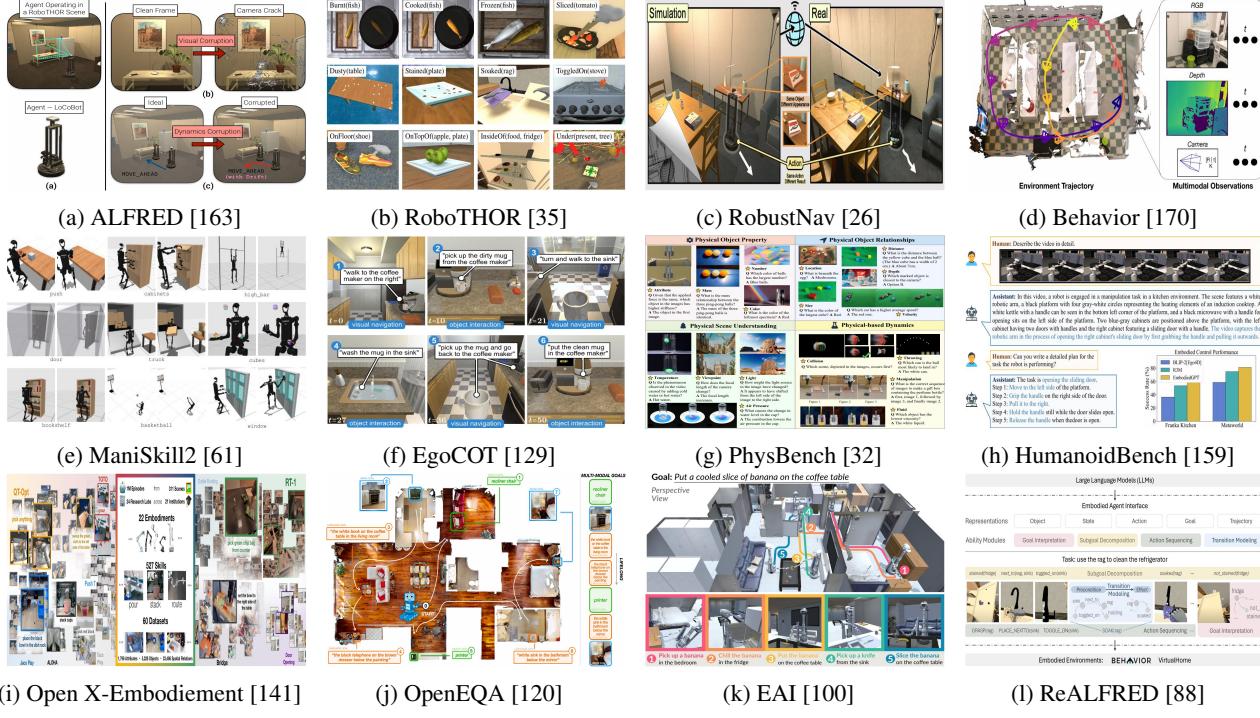


Figure 14: An overview of all multi-embodied benchmarks listed in Table 2.

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 [93, 96]. 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 [233, 118, 218, 245, 174], as shown in Figure 13(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 [253, 151, 127]. 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 [73, 82, 252, 254, 123, 199], as shown in Figure 13(e). In higher-level language-abstracted environments, some methods directly use pre-trained generative models as world models for future trajectory predictions [134]. 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 [204, 197, 225, 242].

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 and present them in Figure 14. The listed benchmarks include:

- ALFRED [163] 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 in-

Table 2: Benchmarks for testing single agent embodied AI.

	Year	Category	Input				View	Data	Agent Type								
			Perception	Planning	Control	Image			Global	Local	Real	Sim	Hybrid	Humanoid	Quadruped	UAV/UGV/USV	Manipulator
ALFRED [163]	2020		✓	✓	✓	✓			✓		✓					✓	
RoboTHOR [35]	2020		✓	✓	✓	✓			✓							✓	
RobustNav [26]	2021		✓	✓	✓	✓			✓							✓	
Behavior [170]	2021		✓	✓	✓	✓			✓							✓	
ManiSkill2 [61]	2023		✓	✓	✓	✓	✓	✓	✓							✓	
EgoCOT [129]	2023		✓	✓	✓	✓			✓							✓	
HumanoidBench [159]	2024		✓	✓	✓	✓			✓							✓	
Open X-Embodiment [141]	2024		✓	✓	✓	✓	✓	✓	✓		✓					✓	
OpenEQA [120]	2024		✓	✓	✓	✓	✓	✓	✓		✓					✓	
Embodied-Agent-Interface [100]	2024		✓	✓	✓	✓			✓							✓	
ReALFRED [88]	2024		✓	✓	✓	✓			✓							✓	✓
PhysBench [32]	2025		✓			✓			✓								✓

volving 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 25,743 human-authored instructions paired with 8,055 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 [35] 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 are 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 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 [26] is a benchmark framework 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, multi-modal sensing, and adaptive policy learning.
- Behavior [170] 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 [61] 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 2,000 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 [214] and interfaced via OpenAI Gym [16], 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 [129] provides 2,900 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 [159] 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 [182], 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 [141] 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 [148], 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 [120] is the first open-vocabulary benchmark for embodied question answering, featuring over 1,600 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 [100] 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 [88] 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 30,696 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 [32] is a large benchmark designed for evaluating VLMs on their understanding of physical concepts in real-world scenarios. It consists of 10,002 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. MAS research primarily focuses on enabling effective collaboration. Here, we review recent advances in

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

	Year	Category	Input	View	Agent Type	Implement
		Perception Planning Control	Image Pointcloud Language Proprioception	Global Local	Humanoid Quadruped/Bipedal UAV/UGV/UUV Manipulator No certain type	Solver-based NN-based GM-based
ACE [230]	2023	✓ ✓	✓	✓	✓	✓
HetGPPO [10]	2023	✓ ✓ ✓	✓	✓ ✓	✓	✓
Self-play [55]	2023	✓ ✓ ✓ ✓	✓	✓ ✓	✓	✓
TwoStep [166]	2024	✓ ✓ ✓	✓	✓ ✓	✓	✓
ZeroCAP [188]	2024	✓ ✓ ✓	✓	✓ ✓	✓	✓
COHERENT [109]	2024	✓ ✓ ✓	✓	✓ ✓	✓	✓
LLaMAC [246]	2024	✓ ✓ ✓ ✓	✓	✓ ✓	✓	✓
FaGeL [106]	2024	✓ ✓ ✓	✓	✓ ✓	✓	✓
MacNet [145]	2025	✓ ✓ ✓	✓	✓ ✓	✓	✓
CoELA [248]	2025	✓ ✓ ✓ ✓	✓	✓ ✓	✓	✓
FISER [193]	2025	✓ ✓ ✓ ✓	✓	✓ ✓	✓	✓
REMAC [240]	2025	✓ ✓ ✓ ✓	✓	✓ ✓	✓	✓
LDPD [107]	2025	✓ ✓ ✓	✓	✓ ✓	✓	✓

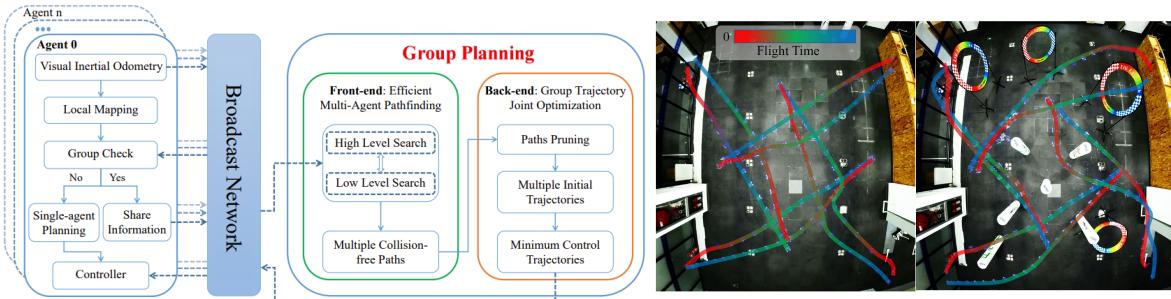


Figure 15: An illustration of control-based multi-agent trajectory optimization [72].

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.

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 [164, 5] 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 [183], making them more suitable for large-scale MAS. Nevertheless, such fully decentralized methods often struggle with resolving conflicts between agents. To overcome these limitations, EMAPF [72] 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, as shown in Figure 15. This enables efficient coordination among large teams of aerial robots. Alternatively, Swarm-Formation [147] 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.

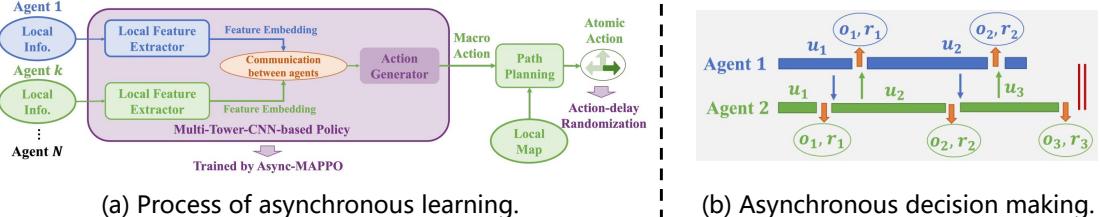


Figure 16: An illustration of multi-agent asynchronous learning [230].

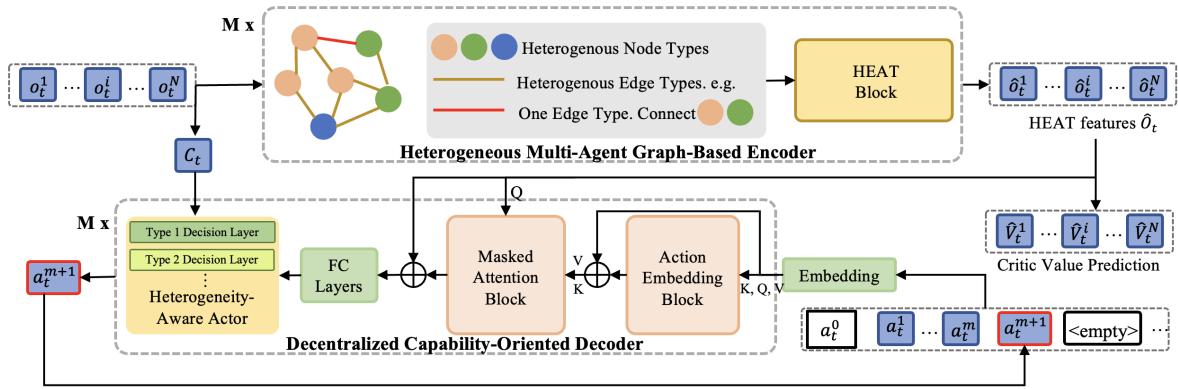


Figure 17: An illustration of multi-agent heterogeneous learning [19].

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 [230] 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, as shown in Figure 16. This approach has proven effective and has inspired several other works tackling asynchronous decision-making challenges [216, 217].

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 HetGPO [10] and COMAT [19] propose using separate observation and policy networks for different agent types. These networks are connected through graph-based communication, enabling effective information exchange, as illustrated in the architecture of COMAT in Figure 17. 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 [262].

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,

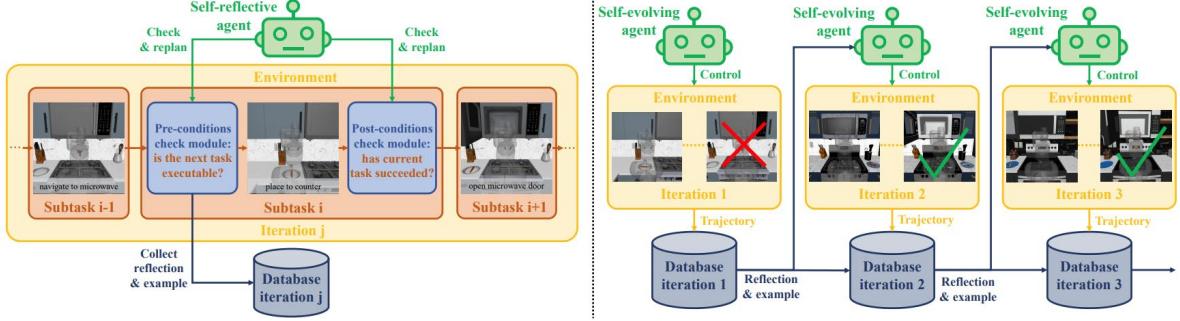


Figure 18: An illustration of self-evolving multi-agent learning [240].

actions, and reward functions), and collaboration patterns (e.g., teammates and opponents) evolve continuously and unpredictably [238]. To address these challenges in MARL, researchers have proposed methods such as robust training [236, 239] and continual coordination [237], 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, as shown in Figure 18. By combining these architectures with distributed policy networks, agents can adapt seamlessly to fluctuations in team size, ensuring robust coordination [19]. 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 [55, 240].

4.3 Generative Models 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.

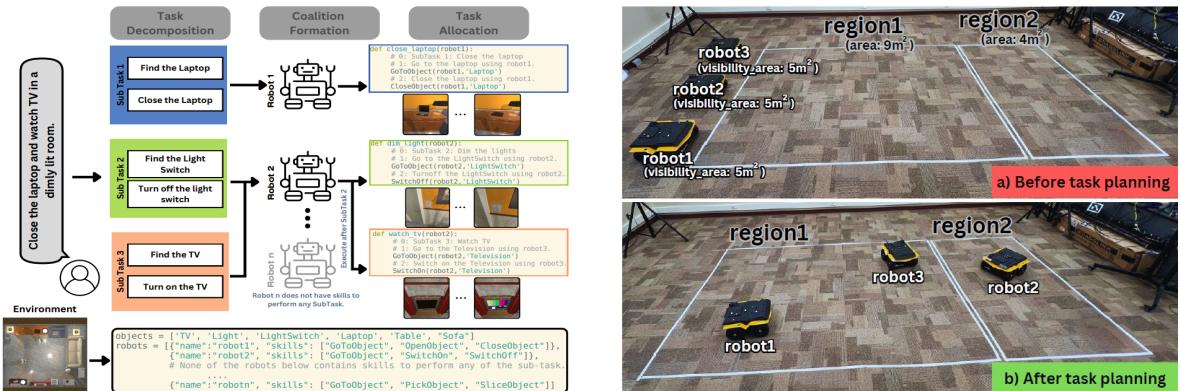


Figure 19: An illustration of generative models based multi-agent task allocation [86].

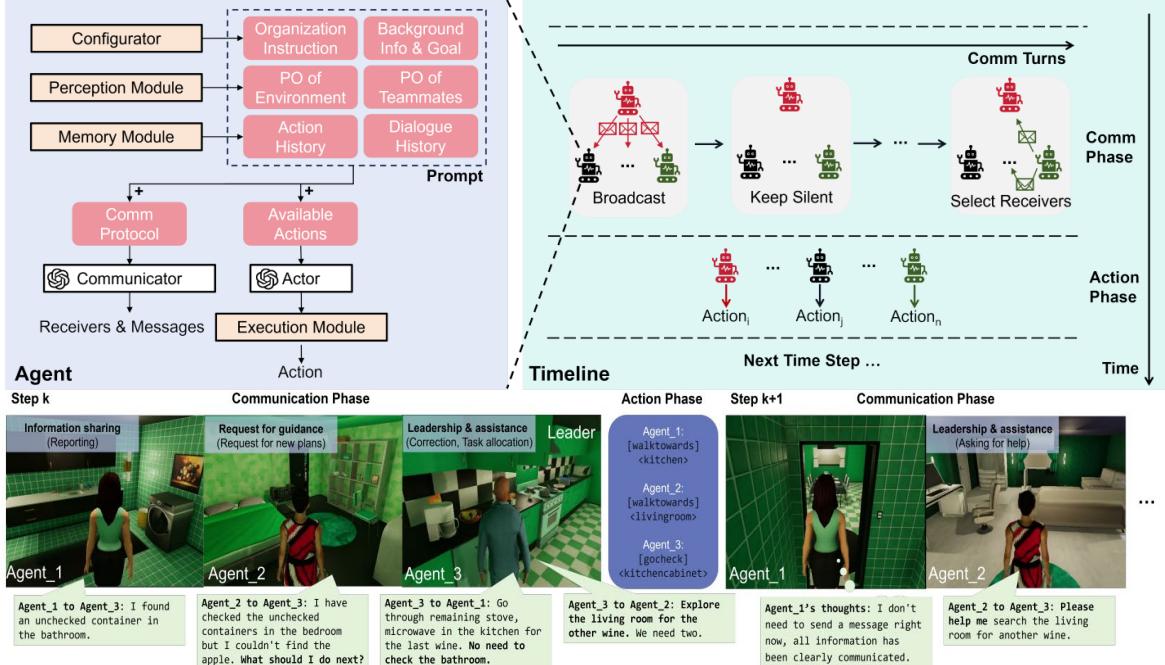


Figure 20: An illustration of generative models based multi-agent distributed collaboration [64].

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 [86] 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 executions. For instance, different robotic vacuum cleaners are each allocated a distinct cleaning area, as illustrated in Figure 19. This paradigm of task decomposition and assignment has become a mainstream planning strategy in embodied MAS [166, 30, 192]. Building on this, a series of works 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 [188, 109, 213, 28]. 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 works have been focused on exploring the use of sub-task dependency graphs to enhance task allocation [205, 135]. 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 [107].

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 [255, 200, 13, 64, 177, 195, 234]. For example, [64] equips 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, as illustrated in Figure 20. However, challenges 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, enhanc-



Figure 21: An illustration of generative models based human-AI collaboration [248].

ing their overall decision-making capabilities [255, 64]. 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. [13]. 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 [246, 145].

Beyond classic centralized or decentralized frameworks for embodied collaboration, generative models introduce a unique advantage: they allow negotiation among agents [113, 27, 64]. 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 [27]. 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 [113, 64].

Human-AI coordination. Efficient collaboration between humans and embodied agents has long been a critical objective of research [1]. Human-AI coordination [190], closely related to research fields such as human-AI interaction (HAI) [185] and human-robot interaction (HRI) [137], 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 multi-modal 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 [248, 173, 64, 46, 62, 106, 4, 108, 54] as illustrated in Figure 21.

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 [248, 173, 64]. 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 [46]. 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 [51]. 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 [193]. 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 [106].

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. 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 [91], as shown in Figure 22(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-

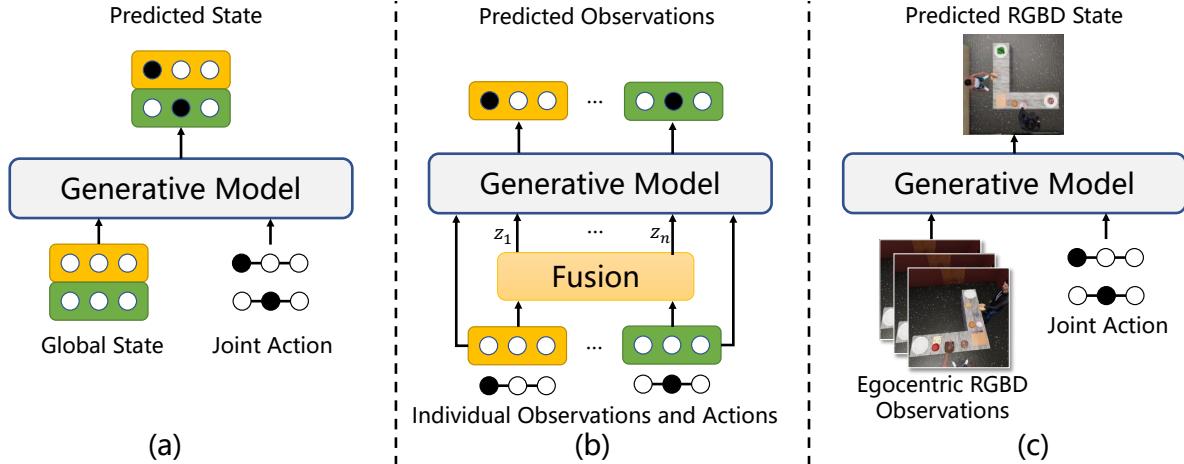


Figure 22: An illustration of different multi-agent world model structures.

regressive trajectory prediction, it is now possible to model the dynamic collaboration of MAS using world models [162, 189, 6, 112], as shown in Figure 22(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 [256]. 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 [249], as shown in Figure 22(c).

4.4 Benchmarks

Table 4: Benchmarks for testing multi-agent embodied AI.

	Year	Category	Input	View	Data	Agent Type
		Perception Planning Control	Image Pointcloud Language Proprioception	Global Local	Real Sim Hybrid	Humanoid/Human Quadruped UAV/UGV/USV Manipulator No certain type
Grid-MAPF [171]	2019				✓	
FurnMove [80]	2020	✓	✓		✓	
SMARTS [263]	2021	✓	✓		✓	
Megaverse [143]	2021	✓	✓		✓	
V2X-Sim [101]	2022	✓			✓	
Bi-DexHands [31]	2022	✓	✓		✓	
MRP-Bench [157]	2023	✓	✓		✓	
LEMMA [58]	2023	✓	✓		✓	
CHAIC [39]	2024	✓	✓		✓	
RoCo [122]	2024	✓	✓		✓	
MAP-THOR [130]	2024	✓	✓		✓	
REMROC [68]	2024	✓	✓		✓	
T2E [247]	2024	✓	✓		✓	
SocialGym 2.0 [24]	2024	✓	✓		✓	
Cambridge RoboMaster [12]	2024	✓	✓		✓	
S3E [45]	2024	✓	✓		✓	
PARTNR [25]	2024	✓	✓		✓	
EmbodiedBench [223]	2025	✓	✓		✓	
MRMG-Planning [65]	2025	✓	✓		✓	
CARIC [22]	2025	✓	✓		✓	

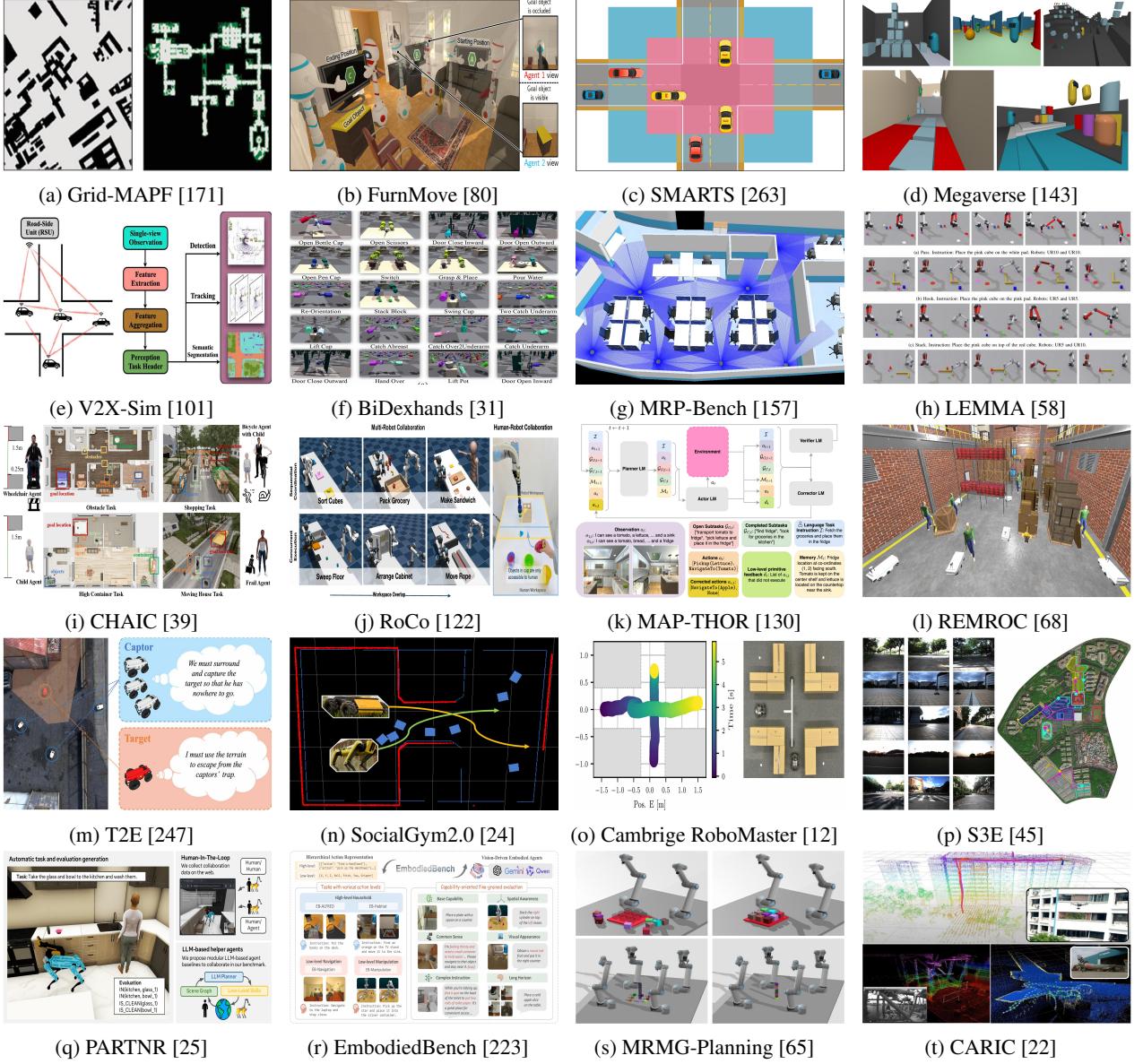


Figure 23: An overview of multi-embodied benchmarks listed in Table 4.

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 paper provides a comprehensive and systematic review of current benchmarks for multi-agent embodied intelligence (see Table 4 and Figure 23). 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 [171] 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 [80] 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 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 [263] 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 [143] 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 1M 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 [101] 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 multi-modal 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 [31] is a unified benchmark built on Isaac Gym [121] that achieves over 30,000 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 [20] and SAPIEN datasets to evaluate task generalization. The platform also provides rich multi-modal 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 [157] 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 [119], Gazebo [90], 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 [58] 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 6,400 expert demonstrations paired with corresponding language commands, along with rich multi-modal 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 [39] 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 [122] 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 [130] 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 [68] 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 [247] 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 [24] 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 [180] 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 [12] 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[9], 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 [45] 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 [25] [25] is a large-scale benchmark for embodied human–AI collaboration driven by natural language. It includes 100,000 procedurally generated tasks distributed across 60 diverse multi-room houses containing 5,819 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 [223] provides a comprehensive suite of 1,128 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, multi-modal evaluation of 19 state-of-the-art proprietary and open-source multi-modal large language models (MLLMs), enabling systematic assessment across a wide range of embodied AI challenges.
- MRMG-Planning [65] formalizes the problem of multi-modal, 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 [22] 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 The Challenges and Future Works

Despite of 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 [251]. Methods such as value decomposition (e.g., VDN [175], QMIX [149]) have improved scalability by decomposing joint objectives into individual utilities, while counterfactual reasoning approaches like COMA [49] 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 [33]. 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 still remain inadequately understood [172]. To address these limitations, future research should explore new theoretical paradigms specifically tailored to embodied MAS, leveraging approaches such as causal infer-

ence 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 [149] with its monotonic value decomposition and MAPPO [229], an adaptation of proximal policy optimization, have demonstrated compelling results in benchmark environments such as SMAC and Google research football (GRF)[238]. Nevertheless, these successes predominantly rely on the CTDE paradigm[3], 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 multi-modal 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 classic 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 exponentially increase complexity, severely reducing sample efficiency and complicating joint exploration due to vast state and policy spaces [2]. These challenges are amplified in real-world contexts, where each interaction incurs considerable time, financial cost, and hardware wear [235]. Although recent advancements in single-agent embodied learning—such as extracting knowledge from large multi-modal models like LLMs [23], employing world-model-based simulations for policy rollouts [116], and leveraging offline data-driven policy calibration [84]—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 multi-modal learning landscapes by offering powerful capabilities in representation, perception, and reasoning [75]. 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 [30], 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 involve high-dimensional multi-modal 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, multi-modal joint pretraining and simulation-driven data augmentation provide valuable transferable knowledge [117] 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 [238]. Single-agent generalist models like Gato [150] and RT-X [141] demonstrate unified architectures' potential 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) [208] 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 that commonly assumed in embodied AI research, scenarios with open environments [266] present significant challenges for embodied MAS due to their inherently dynamic, uncertain and non-stationary characteristics. 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 coordinations. The 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 [238]. 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 coordinations 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. Establishing rigorous and comprehensive evaluation frameworks is crucial yet challenging for advancing multi-agent embodied AI. Recent platforms, including CHAIC [39] (further benchmarks in Table 4), have provided valuable resources but often suffer from limited representational complexity and insufficient 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. Subsequently, future research should aim to develop general-purpose, modular and scalable evaluation frameworks analogous to Gym, MuJoCo or PyMARL [156], incorporating multi-modal interactions, heterogeneous coordination and reproducibility features. Physical testbeds, including robot soccer platforms (e.g., RoboCup [44]) and drone swarm arenas are also essential to validate real-world feasibility and bridge sim-to-real gaps. Moreover, the absence of standardized and interpretable evaluation metrics remains unresolved, as existing measures (e.g., coordination scores, network-based metrics) often lack consistency and generalizability. Therefore, developing unified benchmarks with clearly defined tasks, comprehensive evaluation criteria (robustness, scalability, energy efficiency, behavioral diversity), open-access leader-boards, and formal verification methods is imperative, especially for safety-critical applications.

Applications and implementations. Multi-agent embodied AI holds immense promise across diverse domains including robotics, education, healthcare, military operations, interactive simulations and smart-city infrastructures. In robotics, particularly within industrial manufacturing, warehouse logistics, and autonomous driving, successful deployment necessitates robust real-time coordination, reliable collision avoidance, and adaptable platforms, emphasizing algorithmic generalizability. Educational applications involving virtual agent teams demand controlled, socially appropriate behaviors, thus placing substantial requirements on advanced natural language processing and affective computing capabilities. In healthcare, integrating robotic agents within medical teams highlights the necessity for reliable decision-making, robust privacy safeguards, and strict adherence to ethical standards. Military applications, such as coordinated drone swarms, further underline these challenges, stressing policy robustness under adversarial conditions and the crucial role of human oversight to mitigate ethical and operational risks associated with autonomy. Interactive simulations and competitive gaming environments, exemplified by platforms like OpenAI Five [8] and AlphaStar [191], have provided valuable benchmarks but face limitations due to high computational requirements and difficulties in transferring learned strategies to real-world contexts. Smart-city scenarios, including traffic management and energy-grid optimization, involve extensive interactions among numerous autonomous agents, necessitating highly reliable, efficient, and safe solutions with direct societal impacts. Addressing these diverse yet interconnected challenges requires integrating expert domain knowledge, rigorous validation procedures, and comprehensive safety frameworks, enabling practical real-world deployments of embodied intelligent MAS.

Other future directions. While the previously sections highlight significant challenges within multi-agent embodied AI, several additional fundamental issues warrant deeper investigation. First, multi-modal 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 [41]. 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 [131]. Lastly, as MAS are increasingly gain autonomy, addressing safety and ethical considerations becomes critically important, especially within sensitive or high-stakes environments [155]. 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 classic 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.

References

- [1] Arash Ajoudani, Andrea Maria Zanchettin, Serena Ivaldi, Alin Albu-Schäffer, Kazuhiro Kosuge, and Oussama Khatib. 2018. Progress and prospects of the human–robot collaboration. *Autonomous robots* 42 (2018), 957–975. 4.3
- [2] Stefano V Albrecht, Filippos Christianos, and Lukas Schäfer. 2024. *Multi-agent reinforcement learning: Foundations and modern approaches*. 1, 2.2, 5
- [3] Christopher Amato. 2024. An introduction to centralized training for decentralized execution in cooperative multi-agent reinforcement learning. *arXiv preprint arXiv:2409.03052* (2024). 5
- [4] Kosi Asuzu, Harjinder Singh, and Moad Idrissi. 2025. Human-robot interaction through joint robot planning with large language models. *Intelligent Service Robotics* (2025), 1–17. 4.3
- [5] Federico Augugliaro, Angela P. Schoellig, and Raffaello D’Andrea. 2012. Generation of collision-free trajectories for a quadrocopter fleet: A sequential convex programming approach. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 1917–1922. 4.1
- [6] Paul Barde, Jakob Foerster, Derek Nowrouzezahrai, and Amy Zhang. 2024. 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*. 141–150. 4.3
- [7] Suneel Belkhale, Tianli Ding, Ted Xiao, Pierre Seranet, Quon Vuong, Jonathan Tompson, Yevgen Chebotar, Debidatta Dwibedi, and Dorsa Sadigh. 2024. Rt-h: Action hierarchies using language. *arXiv preprint arXiv:2403.01823* (2024). 1
- [8] Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung, Przemysław Dębiak, Christy Dennison, David Farhi, Quirin Fischer, Shariq Hashme, Chris Hesse, et al. 2019. Dota 2 with large scale deep reinforcement learning. *arXiv preprint arXiv:1912.06680* (2019). 5
- [9] Matteo Bettini, Ryan Kortvelesy, Jan Blumenkamp, and Amanda Prorok. 2024. VMAS: A Vectorized Multi-agent Simulator for Collective Robot Learning. In *Distributed Autonomous Robotic Systems*. 42–56. 4.4
- [10] Matteo Bettini, Ajay Shankar, and Amanda Prorok. 2023. Heterogeneous Multi-Robot Reinforcement Learning. In *AAMAS*. 1485–1494. 3, 4.2

- [11] Johan Bjorck, Fernando Castañeda, Nikita Cherniadev, Xingye Da, Runyu Ding, Linxi Fan, Yu Fang, Dieter Fox, Fengyuan Hu, Spencer Huang, et al. 2025. GR00T N1: An Open Foundation Model for Generalist Humanoid Robots. *arXiv preprint arXiv:2503.14734* (2025). 3.3
- [12] Jan Blumenkamp, Ajay Shankar, Matteo Bettini, Joshua Bird, and Amanda Prorok. 2024. The cambridge robomaster: An agile multi-robot research platform. *arXiv preprint arXiv:2405.02198* (2024). 4, 23o, 4.4
- [13] Xiaohe Bo, Zeyu Zhang, Quanyu Dai, Xueyang Feng, Lei Wang, Rui Li, Xu Chen, and Ji-Rong Wen. 2024. Reflective Multi-Agent Collaboration based on Large Language Models. In *Advances in Neural Information Processing Systems*. 138595–138631. 4.3
- [14] Nico Bohlinger, Grzegorz Czechmanowski, Maciej Piotr Krupka, Piotr Kicki, Krzysztof Walas, Jan Peters, and Davide Tateo. 2024. One Policy to Run Them All: an End-to-end Learning Approach to Multi-Embodiment Locomotion. In *Annual Conference on Robot Learning*. 3.2
- [15] Bruce Bridgeman and Philip Tseng. 2011. Embodied cognition and the perception–action link. *Physics of life reviews* 8, 1 (2011), 73–85. 2.1
- [16] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. 2016. OpenAI gym. *arXiv preprint arXiv:1606.01540* (2016). 3.4
- [17] Rodney Brooks. 1986. A robust layered control system for a mobile robot. *IEEE journal on robotics and automation* 2, 1 (1986), 14–23. 1
- [18] Rodney A Brooks. 1991. New approaches to robotics. *Science* 253, 5025 (1991), 1227–1232. 1, 2.1
- [19] Yuxin Cai, Xiangkun He, Hongliang Guo, Wei-Yun Yau, and Chen Lv. 2024. Transformer-based Multi-Agent Reinforcement Learning for Generalization of Heterogeneous Multi-Robot Cooperation. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 13695–13702. 17, 4.2, 4.2
- [20] Berk Calli, Arjun Singh, Aaron Walsman, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M. Dollar. 2015. The YCB object and Model set: Towards common benchmarks for manipulation research. In *2015 International Conference on Advanced Robotics (ICAR)*. 510–517. 4.4
- [21] Hongye Cao, Fan Feng, Jing Huo, and Yang Gao. 2025. Causal action empowerment for efficient reinforcement learning in embodied agents. *Science China Information Sciences* 68, 5 (2025), 150201. 2.4
- [22] Muqing Cao, Thien-Minh Nguyen, Shenghai Yuan, Andreas Anastasiou, Angelos Zacharia, Savvas Papaioannou, Panayiotis Kolios, Christos G Panayiotou, Marios M Polycarpou, Xinhang Xu, et al. 2025. Cooperative Aerial Robot Inspection Challenge: A Benchmark for Heterogeneous Multi-UAV Planning and Lessons Learned. *arXiv preprint arXiv:2501.06566* (2025). 4, 23t, 4.4
- [23] Yuji Cao, Huan Zhao, Yuheng Cheng, Ting Shu, Yue Chen, Guolong Liu, Gaoqi Liang, Junhua Zhao, Jinyue Yan, and Yun Li. 2024. Survey on large language model-enhanced reinforcement learning: Concept, taxonomy, and methods. *IEEE Transactions on Neural Networks and Learning Systems* (2024). 5
- [24] Rohan Chandra, Zayne Sprague, and Joydeep Biswas. 2024. SOCIALGYM 2.0: Simulator for Multi-Robot Learning and Navigation in Shared Human Spaces. *Proceedings of the AAAI Conference on Artificial Intelligence* 38, 21 (2024), 23778–23780. 4, 23n, 4.4
- [25] Matthew Chang, Gunjan Chhablani, Alexander Clegg, Mikael Dallaire Cote, Ruta Desai, Michal Hlavac, Vladimir Karashchuk, Jacob Krantz, Roozbeh Mottaghi, Priyam Parashar, Siddharth Patki, Ishita Prasad, Xavier Puig, Akshara Rai, Ram Ramrakhya, Daniel Tran, Joanne Truong, John M Turner, Eric Undersander, and Tsung-Yen Yang. 2025. PARTNR: A Benchmark for Planning and Reasoning in Embodied Multi-agent Tasks. In *The Thirteenth International Conference on Learning Representations*. 4, 23q, 4.4
- [26] Prithvijit Chattopadhyay, Judy Hoffman, Roozbeh Mottaghi, and Aniruddha Kembhavi. 2021. RobustNav: Towards Benchmarking Robustness in Embodied Navigation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 15691–15700. 14c, 2, 3.4
- [27] Huaben Chen, Wenkang Ji, Lufeng Xu, and Shiyu Zhao. 2023. Multi-agent consensus seeking via large language models. *arXiv preprint arXiv:2310.20151* (2023). 4.3
- [28] Junting Chen, Checheng Yu, Xunzhe Zhou, Tianqi Xu, Yao Mu, Mengkang Hu, Wenqi Shao, Yikai Wang, Guohao Li, and Lin Shao. 2025. EMOS: Embodiment-aware Heterogeneous Multi-robot Operating System with LLM Agents. In *The Thirteenth International Conference on Learning Representations*. 4.3
- [29] Liang Chen, Yichi Zhang, Shuhuai Ren, Haozhe Zhao, Zefan Cai, Yuchi Wang, Peiyi Wang, Tianyu Liu, and Baobao Chang. 2023. Towards end-to-end embodied decision making via multi-modal large language model: Explorations with GPT4-vision and beyond. *arXiv preprint arXiv:2310.02071* (2023). 3.3

- [30] Weizhe Chen, Sven Koenig, and Bistra Dilkina. 2024. Why solving multi-agent path finding with large language model has not succeeded yet. *arXiv preprint arXiv:2401.03630* (2024). 4, 3, 5
- [31] Yuanpei Chen, Tianhao Wu, Shengjie Wang, Xidong Feng, Jiechuan Jiang, Zongqing Lu, Stephen McAleer, Hao Dong, Song-Chun Zhu, and Yaodong Yang. 2022. Towards Human-Level Bimanual Dexterous Manipulation with Reinforcement Learning. In *Advances in Neural Information Processing Systems*. 5150–5163. 4, 23f, 4.4
- [32] Wei Chow, Jiageng Mao, Boyi Li, Daniel Seita, Vitor Guizilini, and Yue Wang. 2025. PhysBench: Benchmarking and Enhancing Vision-Language Models for Physical World Understanding. *arXiv preprint arXiv:2501.16411* (2025). 14g, 2, 3.4
- [33] Tianshu Chu, Sandeep Chinchali, and Sachin Katti. 2020. Multi-agent Reinforcement Learning for Networked System Control. In *International Conference on Learning Representations*. 5
- [34] Neil T Dantam, Zachary K Kingston, Swarat Chaudhuri, and Lydia E Kavraki. 2018. An incremental constraint-based framework for task and motion planning. *The International Journal of Robotics Research* 37, 10 (2018), 1134–1151. 3.1
- [35] Matt Deitke, Winson Han, Alvaro Herrasti, Aniruddha Kembhavi, Eric Kolve, Roozbeh Mottaghi, Jordi Salvador, Dustin Schwenk, Eli VanderBilt, Matthew Wallingford, Luca Weihs, Mark Yatskar, and Ali Farhadi. 2020. RoboTHOR: An Open Simulation-to-Real Embodied AI Platform. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14b, 2, 3.4
- [36] Zeyu Dong, Yimin Zhu, Yansong Li, Kevin Mahon, and Yu Sun. 2024. Generalizing End-To-End Autonomous Driving In Real-World Environments Using Zero-Shot LLMs. In *8th Annual Conference on Robot Learning*. 3.3
- [37] Ali Dorri, Salil S. Kanhere, and Raja Jurdak. 2018. Multi-Agent Systems: A Survey. *IEEE Access* 6 (2018), 28573–28593. 1, 2.2
- [38] Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch, and Pete Florence. 2023. PaLM-E: An Embodied Multimodal Language Model. In *Proceedings of the 40th International Conference on Machine Learning*. 8469–8488. 1, 3.3
- [39] Weihua Du, Qiushi Lyu, Jiaming Shan, Zhenting Qi, Hongxin Zhang, Sunli Chen, Andi Peng, Tianmin Shu, Kwonjoon Lee, Behzad Dariush, and Chuang Gan. 2024. Constrained Human-AI Cooperation: An Inclusive Embodied Social Intelligence Challenge. In *Advances in neural information processing systems*. 4, 23i, 4.4, 5
- [40] Jiafei Duan, Samson Yu, Hui Li Tan, Hongyuan Zhu, and Cheston Tan. 2022. A survey of embodied ai: From simulators to research tasks. *IEEE Transactions on Emerging Topics in Computational Intelligence* 6, 2 (2022), 230–244. 1
- [41] Jianguo Duan, Liwen Zhuang, Qinglei Zhang, Ying Zhou, and Jiyun Qin. 2024. Multimodal perception-fusion-control and human–robot collaboration in manufacturing: A review. *The International Journal of Advanced Manufacturing Technology* 132, 3 (2024), 1071–1093. 5
- [42] Kunpeng Duan, Shanling Dong, Zhen Fan, Senlin Zhang, Yaqing Shu, and Meiqin Liu. 2025. Multimode trajectory tracking control of Unmanned Surface Vehicles based on LSTM assisted Model Predictive Control. *Ocean Engineering* 328 (2025), 121015. 2.5
- [43] Manfred Eppe, Christian Gumbsch, Matthias Kerzel, Phuong DH Nguyen, Martin V Butz, and Stefan Wermter. 2020. Hierarchical principles of embodied reinforcement learning: A review. *arXiv preprint arXiv:2012.10147* (2020). 1
- [44] Kevin Farias, Pablo Moraes, Igor Nunes, Juan Deniz, Sebastian Barcelona, Hiago Sodre, William Moraes, Monica Rodriguez, Ahilen Mazondo, Vincent Sandin, et al. 2025. RoboCup Rescue 2025 Team Description Paper UruBots. *arXiv preprint arXiv:2504.09778* (2025). 5
- [45] Dapeng Feng, Yuhua Qi, Shipeng Zhong, Zhiqiang Chen, Qiming Chen, Hongbo Chen, Jin Wu, and Jun Ma. 2024. S3E: A Multi-Robot Multimodal Dataset for Collaborative SLAM. *IEEE Robotics and Automation Letters* 9, 12 (2024), 11401–11408. 4, 23p, 4.4
- [46] Xueyang Feng, Zhi-Yuan Chen, Yujia Qin, Yankai Lin, Xu Chen, Zhiyuan Liu, and Ji-Rong Wen. 2024. Large language model-based human-agent collaboration for complex task solving. *arXiv preprint arXiv:2402.12914* (2024). 4.3

- [47] Zhaohan Feng, Jie Chen, Wei Xiao, Jian Sun, Bin Xin, and Gang Wang. 2024. Learning Hybrid Policies for MPC with Application to Drone Flight in Unknown Dynamic Environments. *Unmanned Systems* 12, 02 (2024), 429–441. 2.5, 3.2
- [48] Philipp Foehn, Angel Romero, and Davide Scaramuzza. 2021. Time-optimal planning for quadrotor waypoint flight. *Science Robotics* 6, 56 (2021), eabh1221. 3.1
- [49] Jakob N Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2018. Counterfactual multi-agent policy gradients. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. 2974–2982. 5
- [50] Thomas Fork and Francesco Borrelli. 2023. Euclidean and non-Euclidean trajectory optimization approaches for quadrotor racing. *arXiv preprint arXiv:2309.07262* (2023). 3.1
- [51] Taioli Francesco, Zorzi Edoardo, Franchi Gianni, Castellini Alberto, Farinelli Alessandro, Cristani Marco, and Wang Yiming. 2025. Collaborative Instance Object Navigation: Leveraging Uncertainty-Awareness to Minimize Human-Agent Dialogues. *arXiv preprint arXiv:2412.01250* (2025). 4.3
- [52] Jiawei Fu, Yunlong Song, Yan Wu, Fisher Yu, and Davide Scaramuzza. 2023. Learning Deep Sensorimotor Policies for Vision-Based Autonomous Drone Racing. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 5243–5250. 1, 3.2
- [53] Hao Gao, Shaoyu Chen, Bo Jiang, Bencheng Liao, Yiang Shi, Xiaoyang Guo, Yuechuan Pu, Haoran Yin, Xiangyu Li, Xinbang Zhang, et al. 2025. Rad: Training an end-to-end driving policy via large-scale 3dgs-based reinforcement learning. *arXiv preprint arXiv:2502.13144* (2025). 3.2, 3.3
- [54] Jiawei Gao, Ziqin Wang, Zeqi Xiao, Jingbo Wang, Tai Wang, Jinkun Cao, Xiaolin Hu, Si Liu, Jifeng Dai, and Jiangmiao Pang. 2024. CooHOI: Learning Cooperative Human-Object Interaction with Manipulated Object Dynamics. In *Advances in Neural Information Processing Systems*. 79741–79763. 4.3
- [55] Yuan Gao, Junfeng Chen, Xi Chen, Chongyang Wang, Junjie Hu, Fuqin Deng, and Tin Lun Lam. 2023. Asymmetric Self-Play-Enabled Intelligent Heterogeneous Multirobot Catching System Using Deep Multi-agent Reinforcement Learning. *IEEE Transactions on Robotics* 39, 4 (2023), 2603–2622. 3, 4.2
- [56] Yiran Geng, Boshi An, Haoran Geng, Yuanpei Chen, Yaodong Yang, and Hao Dong. 2023. RLAfford: End-to-End Affordance Learning for Robotic Manipulation. In *IEEE International Conference on Robotics and Automation*. 5880–5886. 3.2
- [57] Atharva Ghotavadekar, František Nekovář, Martin Saska, and Jan Faigl. 2025. Variable Time-Step MPC for Agile Multi-Rotor UAV Interception of Dynamic Targets. *IEEE Robotics and Automation Letters* 10, 2 (2025), 1249–1256. 3.1
- [58] Ran Gong, Xiaofeng Gao, Qiaozhi Gao, Suhaila Shakiah, Govind Thattai, and Gaurav S. Sukhatme. 2023. LEMMA: Learning Language-Conditioned Multi-Robot Manipulation. *IEEE Robotics and Automation Letters* 8, 10 (2023), 6835–6842. 4, 23h, 4.4
- [59] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680. 2.6
- [60] Albert Gu and Tri Dao. 2024. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. In *First Conference on Language Modeling*. 2.7
- [61] Jiayuan Gu, Fanbo Xiang, Xuanlin Li, Zhan Ling, Xiqiang Liu, Tongzhou Mu, Yihe Tang, Stone Tao, Xinyue Wei, Yunchao Yao, Xiaodi Yuan, Pengwei Xie, Zhiao Huang, Rui Chen, and Hao Su. 2023. ManiSkill2: A Unified Benchmark for Generalizable Manipulation Skills. In *The Eleventh International Conference on Learning Representations*. 14e, 2, 3.4
- [62] Cong Guan, Lichao Zhang, Chunpeng Fan, Yichen Li, Feng Chen, Lihe Li, Yunjia Tian, Lei Yuan, and Yang Yu. 2023. Efficient Human-AI Coordination via Preparatory Language-based Convention. *arXiv preprint arXiv:2311.00416* (2023). 4.3
- [63] Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. 2024. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680* (2024). 2.2
- [64] Xudong Guo, Kaixuan Huang, Jiale Liu, Wenhui Fan, Natalia Vélez, Qingyun Wu, Huazheng Wang, Thomas L. Griffiths, and Mengdi Wang. 2024. Embodied LLM Agents Learn to Cooperate in Organized Teams. In *Language Gamification - NeurIPS 2024 Workshop*. 20, 4.3, 4.3

- [65] Valentin N Hartmann, Tirza Heinle, and Stelian Coros. 2025. A Benchmark for Optimal Multi-Modal Multi-Robot Multi-Goal Path Planning with Given Robot Assignment. *arXiv preprint arXiv:2503.03509* (2025). 4, 23s, 4.4
- [66] Alex Havrilla, Yuqing Du, Sharath Chandra Raparthi, Christoforos Nalmpantis, Jane Dwivedi-Yu, Maksym Zhuravinskyi, Eric Hambro, Sainbayar Sukhbaatar, and Roberta Raileanu. 2024. Teaching large language models to reason with reinforcement learning. *arXiv preprint arXiv:2403.04642* (2024). 2.4
- [67] Tairan He, Chong Zhang, Wenli Xiao, Guanqi He, Changliu Liu, and Guanya Shi. 2024. Agile but safe: Learning collision-free high-speed legged locomotion. *arXiv preprint arXiv:2401.17583* (2024). 3.2
- [68] Lukas Heuer, Luigi Palmieri, Anna Mannucci, Sven Koenig, and Martin Magnusson. 2024. Benchmarking Multi-Robot Coordination in Realistic, Unstructured Human-Shared Environments. In *2024 IEEE International Conference on Robotics and Automation*. 14541–14547. 4, 23l, 4.4
- [69] Jonathan Ho and Stefano Ermon. 2016. Generative adversarial imitation learning. In *Advances in neural information processing systems*. 4572–4580. 2.6
- [70] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*. 6840–6851. 2.7
- [71] Chuye Hong, Kangyao Huang, and Huaping Liu. 2024. Learning a Distributed Hierarchical Locomotion Controller for Embodied Cooperation. In *Annual Conference on Robot Learning*. 3.2
- [72] Jialiang Hou, Xin Zhou, Zhongxue Gan, and Fei Gao. 2022. Enhanced Decentralized Autonomous Aerial Robot Teams With Group Planning. *IEEE Robotics and Automation Letters* 7, 4 (2022), 9240–9247. 15, 4.1
- [73] Anthony Hu, Lloyd Russell, Hudson Yeo, Zak Murez, George Fedoseev, Alex Kendall, Jamie Shotton, and Gianluca Corrado. 2023. Gaia-1: A generative world model for autonomous driving. *arXiv preprint arXiv:2309.17080* (2023). 3.3
- [74] Yingdong Hu, Fanqi Lin, Tong Zhang, Li Yi, and Yang Gao. 2023. Look before you leap: Unveiling the power of GPT-4v in robotic vision-language planning. *arXiv preprint arXiv:2311.17842* (2023). 3.3
- [75] Dawei Huang, Chuan Yan, Qing Li, and Xiaojiang Peng. 2024. From Large Language Models to Large Multi-modal Models: A Literature Review. *Applied Sciences* 14, 12 (2024). 5
- [76] Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, Pierre Sermanet, Tomas Jackson, Noah Brown, Linda Luu, Sergey Levine, Karol Hausman, and brian ichter. 2023. Inner Monologue: Embodied Reasoning through Planning with Language Models. In *Proceedings of The 6th Conference on Robot Learning*. 1769–1782. 1, 3.3
- [77] Yidong Huang, Jacob Sansom, Ziqiao Ma, Felix Gervits, and Joyce Chai. 2024. DriVLMe: Enhancing LLM-based Autonomous Driving Agents with Embodied and Social Experiences. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 3153–3160. 3.3
- [78] brian ichter, Anthony Brohan, Yevgen Chebotar, Chelsea Finn, Karol Hausman, Alexander Herzog, Daniel Ho, Julian Ibarz, Alex Irpan, Eric Jang, Ryan Julian, Dmitry Kalashnikov, Sergey Levine, Yao Lu, Carolina Parada, Kanishka Rao, Pierre Sermanet, Alexander T Toshev, Vincent Vanhoucke, Fei Xia, Ted Xiao, Peng Xu, Mengyuan Yan, Noah Brown, Michael Ahn, Omar Cortes, Nicolas Sievers, Clayton Tan, Sichun Xu, Diego Reyes, Jarek Rettinghouse, Jornell Quiambao, Peter Pastor, Linda Luu, Kuang-Huei Lee, Yuheng Kuang, Sally Jesmonth, Nikhil J. Joshi, Kyle Jeffrey, Rosario Jauregui Ruano, Jasmine Hsu, Keerthana Gopalakrishnan, Byron David, Andy Zeng, and Chuyuan Kelly Fu. 2023. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. In *Proceedings of The 6th Conference on Robot Learning*. 287–318. 6, 2.5, 1, 3.3
- [79] Physical Intelligence, Kevin Black, Noah Brown, James Darpinian, Karan Dhabalia, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, et al. 2025. *pi_{(0.5)}*: A Vision-Language-Action Model with Open-World Generalization. *arXiv preprint arXiv:2504.16054* (2025). 1, 3.3
- [80] Unnat Jain, Luca Weihs, Eric Kolve, Ali Farhadi, Svetlana Lazebnik, Aniruddha Kembhavi, and Alexander Schwing. 2020. A Cordial Sync: Going Beyond Marginal Policies for Multi-agent Embodied Tasks. In *Computer Vision – ECCV 2020*. 471–490. 4, 23b, 4.4
- [81] Biao Jia and Dinesh Manocha. 2024. Sim-to-Real Robotic Sketching using Behavior Cloning and Reinforcement Learning. In *2024 IEEE International Conference on Robotics and Automation*. 18272–18278. 12, 3.2
- [82] Fan Jia, Weixin Mao, Yingfei Liu, Yucheng Zhao, Yuqing Wen, Chi Zhang, Xiangyu Zhang, and Tiancai Wang. 2023. Adriver-i: A general world model for autonomous driving. *arXiv preprint arXiv:2311.13549* (2023). 3.3
- [83] Kemou Jiang, Xuan Cai, Zhiyong Cui, Aoyong Li, Yilong Ren, Haiyang Yu, Hao Yang, Daocheng Fu, Licheng Wen, and Pinlong Cai. 2024. KoMA: Knowledge-Driven Multi-Agent Framework for Autonomous Driving With Large Language Models. *IEEE Transactions on Intelligent Vehicles* (2024), 1–15. 3.3

- [84] Tao Jiang, Lei Yuan, Lihe Li, Cong Guan, Zongzhang Zhang, and Yang Yu. 2025. Multi-Agent Domain Calibration with a Handful of Offline Data. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. 5
- [85] Yu Kang, Jian Di, Ming Li, Yunbo Zhao, and Yuhui Wang. 2024. Autonomous multi-drone racing method based on deep reinforcement learning. *Science China Information Sciences* 67, 8 (2024), 180203. 2.4
- [86] Shyam Sundar Kannan, Vishnunandan L. N. Venkatesh, and Byung-Cheol Min. 2024. SMART-LLM: Smart Multi-Agent Robot Task Planning using Large Language Models. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 12140–12147. 19, 4.3
- [87] Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan P Foster, Pannag R Sanketi, Quan Vuong, Thomas Kollar, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, and Chelsea Finn. 2024. OpenVLA: An Open-Source Vision-Language-Action Model. In *8th Annual Conference on Robot Learning*. 1, 3.3
- [88] Taewoong Kim, Cheolhong Min, Byeonghwi Kim, Jinyeon Kim, Wonje Jeung, and Jonghyun Choi. 2025. ReALFRED: An Embodied Instruction Following Benchmark in Photo-Realistic Environments. In *Computer Vision – ECCV 2024*. 346–364. 14l, 2, 3.4
- [89] B Ravi Kiran, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A. Al Sallab, Senthil Yogamani, and Patrick Pérez. 2022. Deep Reinforcement Learning for Autonomous Driving: A Survey. *IEEE Transactions on Intelligent Transportation Systems* 23, 6 (2022), 4909–4926. 2.4
- [90] N. Koenig and A. Howard. 2004. Design and use paradigms for Gazebo, an open-source multi-robot simulator. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol. 3. 2149–2154. 4.4
- [91] Orr Krupnik, Igor Mordatch, and Aviv Tamar. 2020. Multi-Agent Reinforcement Learning with Multi-Step Generative Models. In *Proceedings of the Conference on Robot Learning*, Vol. 100. 776–790. 4.3
- [92] Dana Kulic, Gentiane Venture, Kostas Bekris, and Enrique Coronado. 2024. MPCC++: Model Predictive Contouring Control for Time-Optimal Flight with Safety Constraints. In *Robotics: Science and Systems Conference*. 3.1
- [93] Minae Kwon, Sang Michael Xie, Kalesha Bullard, and Dorsa Sadigh. 2023. Reward Design with Language Models. In *The Eleventh International Conference on Learning Representations*. 3.3
- [94] Guillaume Lample and Devendra Singh Chaplot. 2017. Playing FPS games with deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 2.4
- [95] S. Lavavalle. 1998. Rapidly-exploring random trees: A new tool for path planning. *Research Report 9811* (1998). 1, 3.1
- [96] Yujeong Lee, Sangwoo Shin, Wei-Jin Park, and Honguk Woo. 2024. LLM-Based Offline Learning for Embodied Agents via Consistency-Guided Reward Ensemble. *arXiv preprint arXiv:2411.17135* (2024). 3.3
- [97] Jinlong Li, Xinyu Liu, Baolu Li, Runsheng Xu, Jiachen Li, Hongkai Yu, and Zhengzhong Tu. 2024. Comamba: Real-time cooperative perception unlocked with state space models. *arXiv preprint arXiv:2409.10699* (2024). 3.3
- [98] Jinning Li, Chen Tang, Masayoshi Tomizuka, and Wei Zhan. 2022. Hierarchical Planning Through Goal-Conditioned Offline Reinforcement Learning. *IEEE Robotics and Automation Letters* 7, 4 (2022), 10216–10223. 2.5
- [99] Lingyun Luke Li, Bin Yang, Ming Liang, Wenyuan Zeng, Mengye Ren, Sean Segal, and Raquel Urtasun. 2020. End-to-end Contextual Perception and Prediction with Interaction Transformer. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 5784–5791. 3.3
- [100] Manling Li, Shiyu Zhao, Qineng Wang, Kangrui Wang, Yu Zhou, Sanjana Srivastava, Cem Gokmen, Tony Lee, Li Erran Li, Ruohan Zhang, Weiyu Liu, Percy Liang, Li Fei-Fei, Jiayuan Mao, and Jiajun Wu. 2024. Embodied Agent Interface: Benchmarking LLMs for Embodied Decision Making. In *Advances in Neural Information Processing Systems*. 100428–100534. 14k, 2, 3.4
- [101] Yiming Li, Dekun Ma, Ziyan An, Zixun Wang, Yiqi Zhong, Siheng Chen, and Chen Feng. 2022. V2X-Sim: Multi-Agent Collaborative Perception Dataset and Benchmark for Autonomous Driving. *IEEE Robotics and Automation Letters* 7, 4 (2022), 10914–10921. 4, 23e, 4.4
- [102] Yifei Li, Xin Wang, Jian Sun, Gang Wang, and Jie Chen. 2023. Data-driven consensus control of fully distributed event-triggered multi-agent systems. *Science China Information Sciences* 66, 5 (2023), 152202. 3

- [103] Zhuo Li, Weiran Wu, Jialin Wang, Gang Wang, and Jian Sun. 2024. Continuous advantage learning for minimum-time trajectory planning of autonomous vehicles. *Science China Information Sciences* 67, 7 (2024), 172206. 2.4
- [104] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2015. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971* (2015). 2.4
- [105] Bill Yuchen Lin, Yicheng Fu, Karina Yang, Faeze Brahman, Shiyu Huang, Chandra Bhagavatula, Prithviraj Ammanabrolu, Yejin Choi, and Xiang Ren. 2023. SwiftSage: A Generative Agent with Fast and Slow Thinking for Complex Interactive Tasks. In *Advances in Neural Information Processing Systems*. 23813–23825. 3.3
- [106] Jia Liu and Min Chen. 2024. FaGeL: Fabric LLMs Agent empowered Embodied Intelligence Evolution with Autonomous Human-Machine Collaboration. *arXiv preprint arXiv:2412.20297* (2024). 3, 4.3
- [107] Jiaqi Liu, Chengkai Xu, Peng Hang, Jian Sun, Mingyu Ding, Wei Zhan, and Masayoshi Tomizuka. 2025. Language-Driven Policy Distillation for Cooperative Driving in Multi-Agent Reinforcement Learning. *IEEE Robotics and Automation Letters* 10, 5 (2025), 4292–4299. 3, 4.3
- [108] Jie Liu, Pan Zhou, Yingjun Du, Ah-Hwee Tan, Cees GM Snoek, Jan-Jakob Sonke, and Efstratios Gavves. 2024. CaPo: Cooperative Plan Optimization for Efficient Embodied Multi-Agent Cooperation. *arXiv preprint arXiv:2411.04679* (2024). 4.3
- [109] Kehui Liu, Zixin Tang, Dong Wang, Zhigang Wang, Xuelong Li, and Bin Zhao. 2024. Coherent: Collaboration of heterogeneous multi-robot system with large language models. *arXiv preprint arXiv:2409.15146* (2024). 3, 4.3
- [110] Wenjie Liu, Jian Sun, Gang Wang, Francesco Bullo, and Jie Chen. 2023. Data-Driven Resilient Predictive Control Under Denial-of-Service. *IEEE Trans. Automat. Control* 68, 8 (2023), 4722–4737. 3
- [111] Yang Liu, Weixing Chen, Yongjie Bai, Xiaodan Liang, Guanbin Li, Wen Gao, and Liang Lin. 2024. Aligning cyber space with physical world: A comprehensive survey on embodied ai. *arXiv preprint arXiv:2407.06886* (2024). 1, 1, 2.1
- [112] Zeyang Liu, Xinrui Yang, Shiguang Sun, Long Qian, Lipeng Wan, Xingyu Chen, and Xuguang Lan. 2024. Grounded Answers for Multi-agent Decision-making Problem through Generative World Model. In *Advances in Neural Information Processing Systems*. 46622–46652. 4.3
- [113] Zijun Liu, Yanzhe Zhang, Peng Li, Yang Liu, and Diyi Yang. 2023. Dynamic llm-agent network: An llm-agent collaboration framework with agent team optimization. *arXiv preprint arXiv:2310.02170* (2023). 4.3
- [114] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in neural information processing systems*. 6382–6393. 2.4
- [115] Tomás Lozano-Pérez and Leslie Pack Kaelbling. 2014. A constraint-based method for solving sequential manipulation planning problems. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 3684–3691. 3.1
- [116] Fan-Ming Luo, Tian Xu, Hang Lai, Xiong-Hui Chen, Weinan Zhang, and Yang Yu. 2024. A survey on model-based reinforcement learning. *Science China Information Sciences* 67, 2 (2024), 121101. 5
- [117] Yueen Ma, Zixing Song, Yuzheng Zhuang, Jianye Hao, and Irwin King. 2024. A survey on vision-language-action models for embodied ai. *arXiv preprint arXiv:2405.14093* (2024). 1, 2.1, 5
- [118] Yecheng Jason Ma, William Liang, Guanzhi Wang, De-An Huang, Osbert Bastani, Dinesh Jayaraman, Yuke Zhu, Linxi Fan, and Anima Anandkumar. 2023. Eureka: Human-level reward design via coding large language models. *arXiv preprint arXiv:2310.12931* (2023). 3.3
- [119] Steven Macenski, Tully Foote, Brian Gerkey, Chris Lalancette, and William Woodall. 2022. Robot Operating System 2: Design, architecture, and uses in the wild. *Science Robotics* 7, 66 (2022), eabm6074. 4.4
- [120] Arjun Majumdar, Anurag Ajay, Xiaohan Zhang, Pranav Putta, Sriram Yenamandra, Mikael Henaff, Sneha Silwal, Paul Mcvay, Oleksandr Maksymets, Sergio Arnaud, Karmesh Yadav, Qiyang Li, Ben Newman, Mohit Sharma, Vincent Berges, Shiqi Zhang, Pulkit Agrawal, Yonatan Bisk, Dhruv Batra, Mrinal Kalakrishnan, Franziska Meier, Chris Paxton, Alexander Sax, and Aravind Rajeswaran. 2024. OpenEQA: Embodied Question Answering in the Era of Foundation Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16488–16498. 14j, 2, 3.4
- [121] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, and Gavriel State. 2021. Isaac Gym: High Performance GPU Based Physics Simulation For Robot Learning. In *Advances in neural information processing systems*. 4.4

- [122] Zhao Mandi, Shreeya Jain, and Shuran Song. 2024. RoCo: Dialectic Multi-Robot Collaboration with Large Language Models. In *2024 IEEE International Conference on Robotics and Automation*. 286–299. 4, 23j, 4.4
- [123] Pietro Mazzaglia, Tim Verbelen, Bart Dhoedt, A Courville, and S Rajeswar. 2024. GenRL: Multimodal-foundation world models for generalization in embodied agents. In *Advances in neural information processing systems*. 3.3
- [124] Oier Mees, Dibya Ghosh, Karl Pertsch, Kevin Black, Homer Rich Walke, Sudeep Dasari, Joey Hejna, Tobias Kreiman, Charles Xu, Jianlan Luo, You Liang Tan, Dorsa Sadigh, Chelsea Finn, and Sergey Levine. 2024. Octo: An Open-Source Generalist Robot Policy. In *First Workshop on Vision-Language Models for Navigation and Manipulation at ICRA 2024*. 1, 3.3
- [125] Yunpeng Mei, Jian Sun, Zhihong Peng, Fang Deng, Gang Wang, and Jie Chen. 2025. RoG-SAM: A Language-Driven Framework for Instance-Level Robotic Grasping Detection. *IEEE Transactions on Multimedia* (2025), 1–13. 3.3
- [126] Daniel Mellinger and Vijay Kumar. 2011. Minimum snap trajectory generation and control for quadrotors. In *IEEE International Conference on Robotics and Automation*. 2520–2525. 3.1
- [127] Vincent Micheli, Eloi Alonso, and François Fleuret. 2023. Transformers are Sample-Efficient World Models. In *The Eleventh International Conference on Learning Representations*. 3.3
- [128] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. Playing Atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602* (2013). 2.4
- [129] Yao Mu, Qinglong Zhang, Mengkang Hu, Wenhui Wang, Mingyu Ding, Jun Jin, Bin Wang, Jifeng Dai, Yu Qiao, and Ping Luo. 2023. EmbodiedGPT: Vision-Language Pre-Training via Embodied Chain of Thought. In *Advances in Neural Information Processing Systems*. 25081–25094. 14f, 2, 3.4
- [130] Siddharth Nayak, Adelmo Morrison Orozco, Marina Ten Have, Vittal Thirumalai, Jackson Zhang, Darren Chen, Aditya Kapoor, Eric Robinson, Karthik Gopalakrishnan, James Harrison, Brian Ichter, Anuj Mahajan, and Hamsa Balakrishnan. 2024. Long-Horizon Planning for Multi-Agent Robots in Partially Observable Environments. In *Advances in Neural Information Processing Systems*. 67929–67967. 4, 23k, 4.4
- [131] Kamal K Ndousse, Douglas Eck, Sergey Levine, and Natasha Jaques. 2021. Emergent social learning via multi-agent reinforcement learning. In *International conference on machine learning*. 7991–8004. 5
- [132] Allen Newell and Herbert A Simon. 1976. Computer science as empirical inquiry: Symbols and search. *Commun. ACM* 19, 3 (1976), 113–126. 1
- [133] Rui Nian, Jinfeng Liu, and Biao Huang. 2020. A review on reinforcement learning: Introduction and applications in industrial process control. *Computers & Chemical Engineering* 139 (2020), 106886. 2.4
- [134] Kolby Nottingham, Prithviraj Ammanabrolu, Alane Suhr, Yejin Choi, Hannaneh Hajishirzi, Sameer Singh, and Roy Fox. 2023. Do embodied agents dream of pixelated sheep: Embodied decision making using language guided world modelling. In *International Conference on Machine Learning*. 26311–26325. 3.3
- [135] Kazuma Obata, Tatsuya Aoki, Takato Horii, Tadahiro Taniguchi, and Takayuki Nagai. 2025. LiP-LLM: Integrating Linear Programming and Dependency Graph With Large Language Models for Multi-Robot Task Planning. *IEEE Robotics and Automation Letters* 10 (2025), 1122–1129. 4.3
- [136] Hanbit Oh and Takamitsu Matsubara. 2024. Leveraging Demonstrator-Perceived Precision for Safe Interactive Imitation Learning of Clearance-Limited Tasks. *IEEE Robotics and Automation Letters* 9, 4 (2024), 3387–3394. 3.2
- [137] Linda Onnasch and Eileen Roesler. 2021. A taxonomy to structure and analyze human–robot interaction. *International Journal of Social Robotics* 13, 4 (2021), 833–849. 4.3
- [138] Afshin Oroojlooy and Davood Hajinezhad. 2023. A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence* 53, 11 (2023), 13677–13722. 1
- [139] Joaquim Ortiz-Haro, Wolfgang Höning, Valentin N. Hartmann, Marc Toussaint, and Ludovic Righetti. 2024. iDb-RRT: Sampling-based Kinodynamic Motion Planning with Motion Primitives and Trajectory Optimization. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 10702–10709. 3.1
- [140] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Gray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*. 2.4

- [141] Abby O'Neill, Abdul Rehman, Abhiram Maddukuri, Abhishek Gupta, Abhishek Padalkar, Abraham Lee, Acorn Pooley, Agrim Gupta, Ajay Mandlekar, Ajinkya Jain, Albert Tung, Alex Bewley, Alex Herzog, Alex Irpan, Alexander Khazatsky, Anant Rai, Anchit Gupta, Andrew Wang, Anikait Singh, Animesh Garg, Aniruddha Kembhavi, Annie Xie, Anthony Brohan, Antonin Raffin, Archit Sharma, Arefeh Yavary, Arhan Jain, Ashwin Balakrishna, Ayzaan Wahid, Ben Burgess-Limerick, Beomjoon Kim, Bernhard Schölkopf, Blake Wulfe, Brian Ichter, Cewu Lu, Charles Xu, Charlotte Le, Chelsea Finn, Chen Wang, Chenfeng Xu, Cheng Chi, Chenguang Huang, Christine Chan, Christopher Agia, Chuer Pan, Chuyuan Fu, Coline Devin, Danfei Xu, Daniel Morton, Danny Driess, Daphne Chen, Deepak Pathak, Dhruv Shah, Dieter Büchler, Dinesh Jayaraman, Dmitry Kalashnikov, Dorsa Sadigh, Edward Johns, Ethan Foster, Fangchen Liu, Federico Ceola, Fei Xia, Feiyu Zhao, Freek Stulp, Gaoyue Zhou, Gaurav S. Sukhatme, Gautam Salhotra, Ge Yan, Gilbert Feng, Giulio Schiavi, Glen Berseth, Gregory Kahn, Guanzhi Wang, Hao Su, Hao-Shu Fang, Haochen Shi, Henghui Bao, Heni Ben Amor, Henrik I Christensen, Hiroki Furuta, Homer Walke, Hongjie Fang, Huy Ha, Igor Mordatch, Ilija Radosavovic, Isabel Leal, Jacky Liang, Jad Abou-Chakra, Jaehyun Kim, Jaimyn Drake, Jan Peters, Jan Schneider, Jasmine Hsu, Jeannette Bohg, Jeffrey Bingham, Jeffrey Wu, Jensen Gao, Jiaheng Hu, Jiajun Wu, Jialin Wu, Jiankai Sun, Jianlan Luo, Jiayuan Gu, Jie Tan, Jihoon Oh, Jimmy Wu, Jingpei Lu, Jingyun Yang, Jitendra Malik, João Silvério, Joey Hejna, Jonathan Booher, Jonathan Tompson, Jonathan Yang, Jordi Salvador, Joseph J. Lim, Junhyek Han, Kaiyuan Wang, Kanishka Rao, Karl Pertsch, Karol Hausman, Keegan Go, Keerthana Gopalakrishnan, Ken Goldberg, Kendra Byrne, Kenneth Oslund, Kento Kawaharazuka, Kevin Black, Kevin Lin, Kevin Zhang, Kiana Ehsani, Kiran Lekkala, Kirsty Ellis, Krishan Rana, Krishnan Srinivasan, Kuan Fang, Kunal Pratap Singh, Kuo-Hao Zeng, Kyle Hatch, Kyle Hsu, Laurent Itti, Lawrence Yunliang Chen, Lerrel Pinto, Li Fei-Fei, Liam Tan, Linxi Jim Fan, Lionel Ott, Lisa Lee, Luca Wehrs, Magnum Chen, Marion Lepert, Marius Memmel, Masayoshi Tomizuka, Masha Itkina, Mateo Guaman Castro, Max Spero, Maximilian Du, Michael Ahn, Michael C. Yip, Mingtong Zhang, Mingyu Ding, Minho Heo, Mohan Kumar Srirama, Mohit Sharma, Moo Jin Kim, Naoaki Kanazawa, Nicklas Hansen, Nicolas Heess, Nikhil J Joshi, Niko Suenderhauf, Ning Liu, Norman Di Palo, Nur Muhammad Mahi Shafullah, Oier Mees, Oliver Kroemer, Osbert Bastani, Pannag R Sanketi, Patrick Tree Miller, Patrick Yin, Paul Wohlhart, Peng Xu, Peter David Fagan, Peter Mitrano, Pierre Sermanet, Pieter Abbeel, Priya Sundaresan, Qiuyu Chen, Quan Vuong, Rafael Rafailov, Ran Tian, Ria Doshi, Roberto Martín-Martín, Rohan Baijal, Rosario Scalise, Rose Hendrix, Roy Lin, Runjia Qian, Ruohan Zhang, Russell Mendonca, Rutav Shah, Ryan Hoque, Ryan Julian, Samuel Bustamante, Sean Kirmani, Sergey Levine, Shan Lin, Sherry Moore, Shikhar Bahl, Shivin Dass, Shubham Sonawani, Shuran Song, Sichun Xu, Siddhant Haladar, Siddharth Karamcheti, Simeon Adebola, Simon Guist, Soroush Nasiriany, Stefan Schaaf, Stefan Welker, Stephen Tian, Subramanian Ramamoorthy, Sudeep Dasari, Suneel Belkhale, Sungjae Park, Suraj Nair, Suvir Mirchandani, Takayuki Osa, Tanmay Gupta, Tatsuya Harada, Tatsuya Matsushima, Ted Xiao, Thomas Kollar, Tianhe Yu, Tianli Ding, Todor Davchev, Tony Z. Zhao, Travis Armstrong, Trevor Darrell, Trinity Chung, Vidhi Jain, Vincent Vanhoucke, Wei Zhan, Wenxuan Zhou, Wolfram Burgard, Xi Chen, Xiaolong Wang, Xinghao Zhu, Xinyang Geng, Xiyuan Liu, Xu Liangwei, Xuanlin Li, Yao Lu, Yecheng Jason Ma, Yejin Kim, Yevgen Chebotar, Yifan Zhou, Yifeng Zhu, Yilin Wu, Ying Xu, Yixuan Wang, Yonatan Bisk, Yoonyoung Cho, Youngwoon Lee, Yuchen Cui, Yue Cao, Yueh-Hua Wu, Yujin Tang, Yuke Zhu, Yunchu Zhang, Yunfan Jiang, Yunshuang Li, Yunzhu Li, Yusuke Iwasawa, Yutaka Matsuo, Zehan Ma, Zhuo Xu, Zichen Jeff Cui, Zichen Zhang, and Zipeng Lin. 2024. Open X-Embodiment: Robotic Learning Datasets and RT-X Models: Open X-Embodiment Collaboration. In *2024 IEEE International Conference on Robotics and Automation*. 6892–6903. 14i, 2, 3.4, 5
- [142] Shubham Pateria, Budhitama Subagja, Ah-hwee Tan, and Chai Quek. 2021. Hierarchical reinforcement learning: A comprehensive survey. *Comput. Surveys* 54, 5 (2021), 1–35. 2.5
- [143] Aleksei Petrenko, Erik Wijmans, Brennan Shacklett, and Vladlen Koltun. 2021. Megaverse: Simulating Embodied Agents at One Million Experiences per Second. In *Proceedings of the 38th International Conference on Machine Learning*. 8556–8566. 4, 23d, 4.4
- [144] Rolf Pfeifer and Fumiya Iida. 2004. Embodied artificial intelligence: Trends and challenges. *Lecture notes in computer science* (2004), 1–26. 1, 1
- [145] Chen Qian, Zihao Xie, YiFei Wang, Wei Liu, Kunlun Zhu, Hanchen Xia, Yufan Dang, Zhuoyun Du, Weize Chen, Cheng Yang, Zhiyuan Liu, and Maosong Sun. 2025. Scaling Large Language Model-based Multi-Agent Collaboration. In *The Thirteenth International Conference on Learning Representations*. 3, 4.3
- [146] Chao Qin, Maxime SJ Michet, Jingxiang Chen, and Hugh H-T Liu. 2024. Time-optimal gate-traversing planner for autonomous drone racing. In *IEEE International Conference on Robotics and Automation*. 8693–8699. 3.1
- [147] Lun Quan, Longji Yin, Chao Xu, and Fei Gao. 2022. Distributed Swarm Trajectory Optimization for Formation Flight in Dense Environments. In *2022 International Conference on Robotics and Automation*. 4979–4985. 4.1

- [148] Sabela Ramos, Sertan Girgin, Léonard Hussenot, Damien Vincent, Hanna Yakubovich, Daniel Toyama, Anita Gergely, Piotr Stanczyk, Raphael Marinier, Jeremiah Harmsen, et al. 2021. Rlds: An ecosystem to generate, share and use datasets in reinforcement learning. *arXiv preprint arXiv:2111.02767* (2021). 3.4
- [149] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder De Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2020. Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research* 21, 178 (2020), 1–51. 2.4, 5, 5
- [150] Scott Reed, Konrad Zolna, Emilio Parisotto, Sergio Gómez Colmenarejo, Alexander Novikov, Gabriel Barthmaron, Mai Giménez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, Tom Eccles, Jake Bruce, Ali Razavi, Ashley Edwards, Nicolas Heess, Yutian Chen, Raia Hadsell, Oriol Vinyals, Mahyar Bordbar, and Nando de Freitas. 2022. A Generalist Agent. *Transactions on Machine Learning Research* (2022). 5
- [151] Jan Robine, Marc Höftmann, Tobias Uelwer, and Stefan Harmeling. 2023. Transformer-based World Models Are Happy With 100k Interactions. In *The Eleventh International Conference on Learning Representations*. 3.3
- [152] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention*. 234–241. 2.7
- [153] I Michael Ross. 2015. A primer on Pontryagin’s principle in optimal control. (2015). 2.3
- [154] Nicholas Roy, Ingmar Posner, Tim Barfoot, Philippe Beaudoin, Yoshua Bengio, Jeannette Bohg, Oliver Brock, Isabelle Depatie, Dieter Fox, Dan Koditschek, et al. 2021. From machine learning to robotics: Challenges and opportunities for embodied intelligence. *arXiv preprint arXiv:2110.15245* (2021). 1
- [155] Elayne Ruane, Abeba Birhane, and Anthony Ventresque. 2019. Conversational AI: Social and Ethical Considerations. *AICS* 2563 (2019), 104–115. 5
- [156] Mikayel Samvelyan, Tabish Rashid, Christian Schroeder De Witt, Gregory Farquhar, Nantas Nardelli, Tim GJ Rudner, Chia-Man Hung, Philip HS Torr, Jakob Foerster, and Shimon Whiteson. 2019. The starcraft multi-agent challenge. *arXiv preprint arXiv:1902.04043* (2019). 5
- [157] Simon Schaefer, Luigi Palmieri, Lukas Heuer, Ruediger Dillmann, Sven Koenig, and Alexander Kleiner. 2023. A Benchmark for Multi-Robot Planning in Realistic, Complex and Cluttered Environments. In *2023 IEEE International Conference on Robotics and Automation*. 9231–9237. 4, 23g, 4.4
- [158] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017). 2.4
- [159] Carmelo Sferrazza, Dun-Ming Huang, Xingyu Lin, Youngwoon Lee, and Pieter Abbeel. 2024. Humanoid-bench: Simulated humanoid benchmark for whole-body locomotion and manipulation. *arXiv preprint arXiv:2403.10506* (2024). 14h, 2, 3.4
- [160] Hao Shao, Yuxuan Hu, Letian Wang, Guanglu Song, Steven L. Waslander, Yu Liu, and Hongsheng Li. 2024. LMDrive: Closed-Loop End-to-End Driving with Large Language Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15120–15130. 3.3
- [161] Lucy Xiaoyang Shi, Archit Sharma, Tony Z. Zhao, and Chelsea Finn. 2023. Waypoint-Based Imitation Learning for Robotic Manipulation. In *7th Annual Conference on Robot Learning*. 3.2
- [162] Zifeng Shi, Meiqin Liu, Senlin Zhang, Ronghao Zheng, Shanling Dong, and Ping Wei. 2025. GAWM: Global-Aware World Model for Multi-Agent Reinforcement Learning. *arXiv preprint arXiv:2501.10116* (2025). 4.3
- [163] Mohit Shridhar, Jesse Thomason, Daniel Gordon, Yonatan Bisk, Winson Han, Roozbeh Mottaghi, Luke Zettlemoyer, and Dieter Fox. 2020. ALFRED: A Benchmark for Interpreting Grounded Instructions for Everyday Tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14a, 3.4, 2
- [164] David Silver. 2021. Cooperative Pathfinding. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* 1 (2021), 117–122. 4.1
- [165] Bharat Singh, Rajesh Kumar, and Vinay Pratap Singh. 2022. Reinforcement learning in robotic applications: a comprehensive survey. *Artificial Intelligence Review* 55, 2 (2022), 945–990. 2.4
- [166] Ishika Singh, David Traum, and Jesse Thomason. 2024. Twostep: Multi-agent task planning using classical planners and large language models. *arXiv preprint arXiv:2403.17246* (2024). 3, 4.3
- [167] Chan Hee Song, Jiaman Wu, Clayton Washington, Brian M Sadler, Wei-Lun Chao, and Yu Su. 2023. LLM-planner: Few-shot grounded planning for embodied agents with large language models. In *Proceedings of the IEEE/CVF international conference on computer vision*. 2998–3009. 3.3
- [168] Yunlong Song and Davide Scaramuzza. 2022. Policy Search for Model Predictive Control With Application to Agile Drone Flight. *IEEE Transactions on Robotics* 38, 4 (2022), 2114–2130. 2.5, 1, 3.2

- [169] Konstantinos Souchleris, George K Sidiropoulos, and George A Papakostas. 2023. Reinforcement learning in game industry—Review, prospects and challenges. *Applied Sciences* 13, 4 (2023), 2443. 2.4
- [170] Sanjana Srivastava, Chengshu Li, Michael Lingelbach, Roberto Martín-Martín, Fei Xia, Kent Elliott Vainio, Zheng Lian, Cem Gokmen, Shyamal Buch, Karen Liu, Silvio Savarese, Hyowon Gweon, Jiajun Wu, and Li Fei-Fei. 2022. BEHAVIOR: Benchmark for Everyday Household Activities in Virtual, Interactive, and Ecological Environments. In *Proceedings of the 5th Conference on Robot Learning*. 477–490. 14d, 2, 3.4
- [171] Roni Stern, Nathan Sturtevant, Ariel Felner, Sven Koenig, Hang Ma, Thayne Walker, Jiaoyang Li, Dor Atzman, Liron Cohen, TK Kumar, et al. 2019. Multi-agent pathfinding: Definitions, variants, and benchmarks. In *Proceedings of the International Symposium on Combinatorial Search*. 151–158. 4, 23a, 4.4
- [172] James WA Strachan, Dalila Albergo, Giulia Borghini, Oriana Pansardi, Eugenio Scaliti, Saurabh Gupta, Krati Saxena, Alessandro Rufo, Stefano Panzeri, Guido Manzi, et al. 2024. Testing theory of mind in large language models and humans. *Nature Human Behaviour* 8, 7 (2024), 1285–1295. 5
- [173] Nan Sun, Chengming Shi, and Yuwen Dong. 2024. Leveraging LLM-based Multi-Agent Collaboration to Enhance Embodied Reasoning in Human-Populated Environments. In *Submitted to Tsinghua University Course: Advanced Machine Learning & Machine Learning*. 4.3
- [174] Shengjie Sun, Runze Liu, Jiafei Lyu, Jing-Wen Yang, Liangpeng Zhang, and Xiu Li. 2024. A Large Language Model-Driven Reward Design Framework via Dynamic Feedback for Reinforcement Learning. *arXiv preprint arXiv:2410.14660* (2024). 3.3
- [175] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, et al. 2017. Value-decomposition networks for cooperative multi-agent learning. *arXiv preprint arXiv:1706.05296* (2017). 5
- [176] Richard S. Sutton and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. 2.4
- [177] Sinan Tan, Weilai Xiang, Huaping Liu, Di Guo, and Fuchun Sun. 2020. Multi-agent Embodied Question Answering in Interactive Environments. In *Computer Vision – ECCV 2020*. 663–678. 4.3
- [178] Rajarshi Tarafdar. 2025. Embodied and Multi-Agent Reinforcement Learning: Advances, Challenges and Opportunities. *International Journal of Innovative Science and Research Technology* 10, 3 (2025), 3183–3187. 1
- [179] Krystof Teissing, Matej Novosad, Robert Penicka, and Martin Saska. 2024. Real-Time Planning of Minimum-Time Trajectories for Agile UAV Flight. *IEEE Robotics and Automation Letters* 9, 11 (2024), 10351–10358. 3.1
- [180] J Terry, Benjamin Black, Nathaniel Grammel, Mario Jayakumar, Ananth Hari, Ryan Sullivan, Luis S Santos, Clemens Dieffendahl, Caroline Horsch, Rodrigo Perez-Vicente, Niall Williams, Yashas Lokesh, and Praveen Ravi. 2021. PettingZoo: Gym for Multi-Agent Reinforcement Learning. In *Advances in Neural Information Processing Systems*. 15032–15043. 4.4
- [181] Tongxuan Tian, Haoyang Li, Bo Ai, Xiaodi Yuan, Zhiao Huang, and Hao Su. 2025. Diffusion Dynamics Models with Generative State Estimation for Cloth Manipulation. *arXiv preprint arXiv:2503.11999* (2025). 3.3
- [182] Emanuel Todorov, Tom Erez, and Yuval Tassa. 2012. Mujoco: A physics engine for model-based control. In *IEEE/RSJ international conference on intelligent robots and systems*. 5026–5033. 3.4
- [183] Jesus Tordesillas and Jonathan P. How. 2022. MADER: Trajectory Planner in Multiagent and Dynamic Environments. *IEEE Transactions on Robotics* 38, 1 (2022), 463–476. 4.1
- [184] AM Turing. 1950. Computing Machinery and Intelligence. (1950). 1, 2.1
- [185] Niels Van Berkel, Mikael B Skov, and Jesper Kjeldskov. 2021. Human-AI interaction: Intermittent, continuous, and proactive. *Interactions* 28, 6 (2021), 67–71. 4.3
- [186] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*. 2.7
- [187] Sai H. Vemprala, Rogerio Bonatti, Arthur Bucker, and Ashish Kapoor. 2024. ChatGPT for Robotics: Design Principles and Model Abilities. *IEEE Access* 12 (2024), 55682–55696. 3.3
- [188] Vishnunandan LN Venkatesh and Byung-Cheol Min. 2024. ZeroCAP: Zero-Shot Multi-Robot Context Aware Pattern Formation via Large Language Models. *arXiv preprint arXiv:2404.02318* (2024). 3, 4.3
- [189] Aravind Venugopal, Stephanie Milani, Fei Fang, and Balaraman Ravindran. 2024. 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*. 1865–1873. 4.3

- [190] Federico Vicentini. 2021. Collaborative robotics: A survey. *Journal of Mechanical Design* 143, 4 (2021), 040802. 4.3
- [191] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *nature* 575, 7782 (2019), 350–354. 5
- [192] Hanwen Wan, Yifei Chen, Zeyu Wei, Dongrui Li, Zexin Lin, Donghao Wu, Jiu Cheng, Yuxiang Zhang, and Xiaoqiang Ji. 2025. EmbodiedAgent: A Scalable Hierarchical Approach to Overcome Practical Challenge in Multi-Robot Control. *arXiv preprint arXiv:2504.10030* (2025). 4.3
- [193] Yanming Wan, Yue Wu, Yiping Wang, Jiayuan Mao, and Natasha Jaques. 2025. Infer Human’s Intentions Before Following Natural Language Instructions. *Proceedings of the AAAI Conference on Artificial Intelligence* 39, 24 (2025), 25309–25317. 3, 4.3
- [194] Guokang Wang, Yanhong Liu, and Huaping Liu. 2025. One model, two skills: Active vision and action learning model for robotic manipulation. *Science China Information Sciences* 68, 6 (2025), 1–19. 3.2
- [195] Jun Wang, Guocheng He, and Yiannis Kantaros. 2024. Safe task planning for language-instructed multi-robot systems using conformal prediction. *arXiv preprint arXiv:2402.15368* (2024). 4.3
- [196] Lirui Wang, Xinlei Chen, Jialiang Zhao, and Kaiming He. 2024. Scaling Proprioceptive-Visual Learning with Heterogeneous Pre-trained Transformers. In *Advances in Neural Information Processing Systems*. 124420–124450. 3.3
- [197] Lirui Wang, Yiyang Ling, Zhecheng Yuan, Mohit Shridhar, Chen Bao, Yuzhe Qin, Bailin Wang, Huazhe Xu, and Xiaolong Wang. 2023. Gensim: Generating robotic simulation tasks via large language models. *arXiv preprint arXiv:2310.01361* (2023). 3.3
- [198] Pengyu Wang, Jiawei Tang, Hin Wang Lin, Fan Zhang, Chaoqun Wang, Jiankun Wang, Ling Shi, and Max Q.-H. Meng. 2025. MINER-RRT*: A Hierarchical and Fast Trajectory Planning Framework in 3D Cluttered Environments. *IEEE Transactions on Automation Science and Engineering* 22 (2025), 10973–10985. 3.1
- [199] Siyin Wang, Zhaoye Fei, Qinyuan Cheng, Shiduo Zhang, Panpan Cai, Jinlan Fu, and Xipeng Qiu. 2025. World Modeling Makes a Better Planner: Dual Preference Optimization for Embodied Task Planning. *arXiv preprint arXiv:2503.10480* (2025). 3.3
- [200] Weizheng Wang, Ike Obi, and Byung-Cheol Min. 2025. Multi-Agent LLM Actor-Critic Framework for Social Robot Navigation. *arXiv preprint arXiv:2503.09758* (2025). 4.3
- [201] Xiao Wang, Guangyao Chen, Guangwu Qian, Pengcheng Gao, Xiao-Yong Wei, Yaowei Wang, Yonghong Tian, and Wen Gao. 2023. Large-scale multi-modal pre-trained models: A comprehensive survey. *Machine Intelligence Research* 20, 4 (2023), 447–482. 1
- [202] Xin Wang, Jian Sun, Fang Deng, Gang Wang, and Jie Chen. 2023. Event-triggered consensus control of heterogeneous multi-agent systems: Model-and data-based approaches. *Science China Information Sciences* 66, 9 (2023), 192201. 3
- [203] Yujin Wang, Quanfeng Liu, Zhengxin Jiang, Tianyi Wang, Junfeng Jiao, Hongqing Chu, Bingzhao Gao, and Hong Chen. 2025. RAD: Retrieval-Augmented Decision-Making of Meta-Actions with Vision-Language Models in Autonomous Driving. *arXiv preprint arXiv:2503.13861* (2025). 3.3
- [204] Yufei Wang, Zhou Xian, Feng Chen, Tsun-Hsuan Wang, Yian Wang, Katerina Fragkiadaki, Zackory Erickson, David Held, and Chuang Gan. 2023. Robogen: Towards unleashing infinite data for automated robot learning via generative simulation. *arXiv preprint arXiv:2311.01455* (2023). 3.3
- [205] Yongdong Wang, Runze Xiao, Jun Younes Louhi Kasahara, Ryosuke Yajima, Keiji Nagatani, Atsushi Yamashita, and Hajime Asama. 2024. Dart-llm: Dependency-aware multi-robot task decomposition and execution using large language models. *arXiv preprint arXiv:2411.09022* (2024). 4.3
- [206] Zhepei Wang, Xin Zhou, Chao Xu, and Fei Gao. 2022. Geometrically Constrained Trajectory Optimization for Multicopters. *IEEE Transactions on Robotics* 38, 5 (2022), 3259–3278. 1
- [207] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *Advances in Neural Information Processing Systems*, Vol. 35. 24824–24837. 3.3
- [208] Muning Wen, Jakub Kuba, Runji Lin, Weinan Zhang, Ying Wen, Jun Wang, and Yaodong Yang. 2022. Multi-agent reinforcement learning is a sequence modeling problem. In *Advances in Neural Information Processing Systems*. 16509–16521. 5

- [209] Di Wu, Xian Wei, Guang Chen, Hao Shen, Xiangfeng Wang, Wenhao Li, and Bo Jin. 2025. Generative Multi-Agent Collaboration in Embodied AI: A Systematic Review. *arXiv preprint arXiv:2502.11518* (2025). 1
- [210] Feiyang Wu, Zhaoyuan Gu, Hanran Wu, Anqi Wu, and Ye Zhao. 2024. Infer and Adapt: Bipedal Locomotion Reward Learning from Demonstrations via Inverse Reinforcement Learning. In *2024 IEEE International Conference on Robotics and Automation*. 16243–16250. 1
- [211] Jianlong Wu, Wei Sun, Tian Gan, Ning Ding, Feijun Jiang, Jiale Shen, and Liqiang Nie. 2023. Neighbor-Guided Consistent and Contrastive Learning for Semi-Supervised Action Recognition. *IEEE Transactions on Image Processing* 32 (2023), 2215–2227. 3
- [212] Jianlong Wu, Haozhe Yang, Tian Gan, Ning Ding, Feijun Jiang, and Liqiang Nie. 2023. CHMATCH: Contrastive Hierarchical Matching and Robust Adaptive Threshold Boosted Semi-Supervised Learning. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15762–15772. 3
- [213] Yuwei Wu, Yuezhan Tao, Peihan Li, Guangyao Shi, Gaurav S Sukhatmem, Vijay Kumar, and Lifeng Zhou. 2024. Hierarchical LLMs In-the-loop Optimization for Real-time Multi-Robot Target Tracking under Unknown Hazards. *arXiv preprint arXiv:2409.12274* (2024). 4.3
- [214] Fanbo Xiang, Yuzhe Qin, Kaichun Mo, Yikuan Xia, Hao Zhu, Fangchen Liu, Minghua Liu, Hanxiao Jiang, Yifu Yuan, He Wang, Li Yi, Angel X. Chang, Leonidas J. Guibas, and Hao Su. 2020. SAPIEN: A SimulAted Part-Based Interactive ENvironment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3.4
- [215] Wei Xiao, Zhaohan Feng, Ziyu Zhou, Jian Sun, Gang Wang, and Jie Chen. 2025. Time-Optimal Flight in Cluttered Environments via Safe Reinforcement Learning. *Unmanned Systems* 0, 0 (2025), 1–10. 10, 3.2
- [216] Yuchen Xiao, Weihao Tan, and Christopher Amato. 2022. Asynchronous Actor-Critic for Multi-Agent Reinforcement Learning. In *Advances in Neural Information Processing Systems*. 4385–4400. 4.2
- [217] Yuchen Xiao, Weihao Tan, Joshua Hoffman, Tian Xia, and Christopher Amato. 2025. Asynchronous multi-agent deep reinforcement learning under partial observability. *The International Journal of Robotics Research* (2025). 4.2
- [218] Tianbao Xie, Siheng Zhao, Chen Henry Wu, Yitao Liu, Qian Luo, Victor Zhong, Yanchao Yang, and Tao Yu. 2023. Text2reward: Reward shaping with language models for reinforcement learning. *arXiv preprint arXiv:2309.11489* (2023). 3.3
- [219] Yuxi Xie, Kenji Kawaguchi, Yiran Zhao, James Xu Zhao, Min-Yen Kan, Junxian He, and Michael Xie. 2023. Self-Evaluation Guided Beam Search for Reasoning. In *Advances in Neural Information Processing Systems*. 41618–41650. 3.3
- [220] Guangkai Xu, Yongtao Ge, Mingyu Liu, Chengxiang Fan, Kangyang Xie, Zhiyue Zhao, Hao Chen, and Chunhua Shen. 2024. What Matters When Repurposing Diffusion Models for General Dense Perception Tasks? *arXiv preprint arXiv:2403.06090* (2024). 3.3
- [221] Zhiyuan Xu, Kun Wu, Junjie Wen, Jinning Li, Ning Liu, Zhengping Che, and Jian Tang. 2024. A survey on robotics with foundation models: Toward embodied ai. *arXiv preprint arXiv:2402.02385* (2024). 1
- [222] Zhenhua Xu, Yujia Zhang, Enze Xie, Zhen Zhao, Yong Guo, Kwan-Yee K. Wong, Zhenguo Li, and Hengshuang Zhao. 2024. DriveGPT4: Interpretable End-to-End Autonomous Driving Via Large Language Model. *IEEE Robotics and Automation Letters* 9, 10 (2024), 8186–8193. 3.3
- [223] Rui Yang, Hanyang Chen, Junyu Zhang, Mark Zhao, Cheng Qian, Kangrui Wang, Qineng Wang, Teja Venkat Koripella, Marziyeh Movahedi, Manling Li, et al. 2025. EmbodiedBench: Comprehensive Benchmarking Multi-modal Large Language Models for Vision-Driven Embodied Agents. *arXiv preprint arXiv:2502.09560* (2025). 4, 23r, 4.4
- [224] Xintong Yang, Ze Ji, Jing Wu, Yu-Kun Lai, Changyun Wei, Guoliang Liu, and Rossitza Setchi. 2022. Hierarchical Reinforcement Learning With Universal Policies for Multistep Robotic Manipulation. *IEEE Transactions on Neural Networks and Learning Systems* 33, 9 (2022), 4727–4741. 2.5
- [225] Yue Yang, Fan-Yun Sun, Luca Weihs, Eli VanderBilt, Alvaro Herrasti, Winson Han, Jiajun Wu, Nick Haber, Ranjay Krishna, Lingjie Liu, et al. 2024. Holodeck: Language guided generation of 3S embodied AI environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16227–16237. 3.3
- [226] Zihan Yang, Jindou Jia, Yuhang Liu, Kexin Guo, Xiang Yu, and Lei Guo. 2024. TRACE: Trajectory Refinement with Control Error Enables Safe and Accurate Maneuvers. In *IEEE International Conference on Control & Automation*. 154–161. 3.1

- [227] Sheng Yin, Xianghe Pang, Yuanzhuo Ding, Menglan Chen, Yutong Bi, Yichen Xiong, Wenhao Huang, Zhen Xiang, Jing Shao, and Siheng Chen. 2024. SafeAgentBench: A Benchmark for Safe Task Planning of Embodied LLM Agents. *arXiv preprint arXiv:2412.13178* (2024). 3.3
- [228] Chao Yu, Jiming Liu, Shamim Nemati, and Guosheng Yin. 2021. Reinforcement learning in healthcare: A survey. *Comput. Surveys* 55, 1 (2021), 1–36. 2.4
- [229] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2022. The surprising effectiveness of ppo in cooperative multi-agent games. In *Advances in neural information processing systems*. 24611–24624. 2.4, 5
- [230] Chao Yu, Xinyi Yang, Jiaxuan Gao, Jiayu Chen, Yunfei Li, Jijia Liu, Yunfei Xiang, Ruixin Huang, Huazhong Yang, Yi Wu, and Yu Wang. 2023. 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*. 1107–1115. 3, 16, 4.2
- [231] Fujie Yu and Yuan Chen. 2023. Cyl-IRRT*: Homotopy Optimal 3D Path Planning for AUVs by Biasing the Sampling Into a Cylindrical Informed Subset. *IEEE Transactions on Industrial Electronics* 70, 4 (2023), 3985–3994. 3.1
- [232] Huan Yu, Chuanqi Hu, Jin Wang, Guodong Lu, Jie Tu, Zhi Zheng, Jingjing Li, and Fei Gao. 2025. Aggressive Collision-Inclusive Motion Planning. *IEEE/ASME Transactions on Mechatronics* 30, 2 (2025), 1412–1423. 3.1, 9
- [233] Wenhao Yu, Nimrod Gileadi, Chuyuan Fu, Sean Kirmani, Kuang-Huei Lee, Montse Gonzalez Arenas, Hao-Tien Lewis Chiang, Tom Erez, Leonard Hasenclever, Jan Humplik, et al. 2023. Language to rewards for robotic skill synthesis. *arXiv preprint arXiv:2306.08647* (2023). 3.3
- [234] Wenhao Yu, Jie Peng, Yueliang Ying, Sai Li, Jianmin Ji, and Yanyong Zhang. 2024. MHRC: Closed-loop Decentralized Multi-Heterogeneous Robot Collaboration with Large Language Models. *arXiv preprint arXiv:2409.16030* (2024). 4.3
- [235] Yang Yu. 2018. Towards Sample Efficient Reinforcement Learning. In *International Joint Conferences on Artificial Intelligence*. 5739–5743. 5
- [236] Lei Yuan, Tao Jiang, Lihe Li, Feng Chen, Zongzhang Zhang, and Yang Yu. 2024. Robust cooperative multi-agent reinforcement learning via multi-view message certification. *Science China Information Sciences* 67, 4 (2024), 142102. 4.2
- [237] Lei Yuan, Lihe Li, Ziqian Zhang, Fuxiang Zhang, Cong Guan, and Yang Yu. 2024. Multiagent Continual Coordination via Progressive Task Contextualization. *IEEE Transactions on Neural Networks and Learning Systems* (2024). 4.2
- [238] Lei Yuan, Ziqian Zhang, Lihe Li, Cong Guan, and Yang Yu. 2023. A survey of progress on cooperative multi-agent reinforcement learning in open environment. *arXiv preprint arXiv:2312.01058* (2023). 1, 4.2, 5, 5, 5
- [239] Lei Yuan, Ziqian Zhang, Ke Xue, Hao Yin, Feng Chen, Cong Guan, Lihe Li, Chao Qian, and Yang Yu. 2023. Robust multi-agent coordination via evolutionary generation of auxiliary adversarial attackers. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 11753–11762. 4.2
- [240] Puzhen Yuan, Anyuan Ma, Yunchao Yao, Huaxiu Yao, Masayoshi Tomizuka, and Mingyu Ding. 2025. REMAC: Self-Reflective and Self-Evolving Multi-Agent Collaboration for Long-Horizon Robot Manipulation. *arXiv preprint arXiv:2503.22122* (2025). 3, 18, 4.2
- [241] Kevin Zakka, Andy Zeng, Pete Florence, Jonathan Tompson, Jeannette Bohg, and Debidatta Dwibedi. 2022. Xirl: Cross-embodiment inverse reinforcement learning. In *Conference on Robot Learning*. 537–546. 3.2
- [242] Abhay Zala, Jaemin Cho, Han Lin, Jaehong Yoon, and Mohit Bansal. 2024. Enigen: Generating and adapting environments via llms for training embodied agents. *arXiv preprint arXiv:2403.12014* (2024). 3.3
- [243] Maryam Zare, Parham M. Kebria, Abbas Khosravi, and Saeid Nahavandi. 2024. A Survey of Imitation Learning: Algorithms, Recent Developments, and Challenges. *IEEE Transactions on Cybernetics* 54, 12 (2024), 7173–7186. 1, 2.6
- [244] Yanjie Ze, Ge Yan, Yueh-Hua Wu, Annabella Macaluso, Yuying Ge, Jianglong Ye, Nicklas Hansen, Li Erran Li, and Xiaolong Wang. 2023. GNFactor: Multi-Task Real Robot Learning with Generalizable Neural Feature Fields. In *7th Annual Conference on Robot Learning*. 1
- [245] Yuwei Zeng, Yao Mu, and Lin Shao. 2024. Learning reward for robot skills using large language models via self-alignment. *arXiv preprint arXiv:2405.07162* (2024). 3.3

- [246] Bin Zhang, Hangyu Mao, Jingqing Ruan, Ying Wen, Yang Li, Shao Zhang, Zhiwei Xu, Dapeng Li, Ziyue Li, Rui Zhao, Lijuan Li, and Guoliang Fan. 2024. Controlling Large Language Model-based Agents for Large-Scale Decision-Making: An Actor-Critic Approach. In *ICLR 2024 Workshop on Large Language Model (LLM) Agents*. 3, 4.3
- [247] Hao Zhang, Jiaming Chen, Jiyu Cheng, Yibin Li, Simon X. Yang, and Wei Zhang. 2024. Nowhere to Go: Benchmarking Multirobot Collaboration in Target Trapping Environment. *IEEE Transactions on Industrial Electronics* 71, 11 (2024), 14430–14439. 4, 23m, 4.4
- [248] Hongxin Zhang, Weihua Du, Jiaming Shan, Qinhong Zhou, Yilun Du, Joshua B. Tenenbaum, Tianmin Shu, and Chuang Gan. 2024. Building Cooperative Embodied Agents Modularly with Large Language Models. In *The Twelfth International Conference on Learning Representations*. 3, 21, 4.3
- [249] Hongxin Zhang, Zeyuan Wang, Qiushi Lyu, Zheyuan Zhang, Sunli Chen, Tianmin Shu, Behzad Dariush, Kwon-joon Lee, Yilun Du, and Chuang Gan. 2025. COMBO: Compositional World Models for Embodied Multi-Agent Cooperation. In *The Thirteenth International Conference on Learning Representations*. 4.3
- [250] Hangtao Zhang, Chenyu Zhu, Xianlong Wang, Ziqi Zhou, Shengshan Hu, and Leo Yu Zhang. 2024. Badrobot: Jailbreaking LLM-based embodied AI in the physical world. *arXiv preprint arXiv:2407.20242* (2024). 3.3
- [251] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. 2021. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of reinforcement learning and control* (2021), 321–384. 1, 5
- [252] Lunjun Zhang, Yuwen Xiong, Ze Yang, Sergio Casas, Rui Hu, and Raquel Urtasun. 2023. Copilot4d: Learning unsupervised world models for autonomous driving via discrete diffusion. *arXiv preprint arXiv:2311.01017* (2023). 3.3
- [253] Weipu Zhang, Gang Wang, Jian Sun, Yetian Yuan, and Gao Huang. 2023. STORM: Efficient Stochastic Transformer based World Models for Reinforcement Learning. In *Advances in Neural Information Processing Systems*. 27147–27166. 3.3
- [254] Yumeng Zhang, Shi Gong, Kaixin Xiong, Xiaoqing Ye, Xiao Tan, Fan Wang, Jizhou Huang, Hua Wu, and Haifeng Wang. 2024. Bevworld: A multimodal world model for autonomous driving via unified bev latent space. *arXiv preprint arXiv:2407.05679* (2024). 3.3
- [255] Yang Zhang, Shixin Yang, Chenjia Bai, Fei Wu, Xiu Li, Zhen Wang, and Xuelong Li. 2024. Towards efficient LLM-llm grounding for embodied multi-agent collaboration. *arXiv preprint arXiv:2405.14314* (2024). 4.3
- [256] Zhejun Zhang, Alexander Liniger, Dengxin Dai, Fisher Yu, and Luc Van Gool. 2023. TrafficBots: Towards World Models for Autonomous Driving Simulation and Motion Prediction. In *2023 IEEE International Conference on Robotics and Automation*. 1522–1529. 4.3
- [257] Chao Zhao, Jie Xu, Ru Peng, Xingyu Chen, Kuihi Mei, and Xuguang Lan. 2024. Experience Consistency Distillation Continual Reinforcement Learning for Robotic Manipulation Tasks. In *2024 IEEE International Conference on Robotics and Automation*. 501–507. 3
- [258] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A Survey of Large Language Models. *arXiv preprint arXiv:2303.18223* (2023). 2.1
- [259] Zhigen Zhao, Shuo Cheng, Yan Ding, Ziyi Zhou, Shiqi Zhang, Danfei Xu, and Ye Zhao. 2024. A Survey of Optimization-Based Task and Motion Planning: From Classical to Learning Approaches. *IEEE/ASME Transactions on Mechatronics* (2024), 1–27. 3.1
- [260] Sipeng Zheng, Jiazheng Liu, Yicheng Feng, and Zongqing Lu. 2023. Steve-eye: Equipping llm-based embodied agents with visual perception in open worlds. *arXiv preprint arXiv:2310.13255* (2023). 1, 3.3
- [261] Yue Zhong, Jiawen Kang, Jinbo Wen, Dongdong Ye, Jiangtian Nie, Dusit Niyato, Xiaozheng Gao, and Shengli Xie. 2025. Generative Diffusion-Based Contract Design for Efficient AI Twin Migration in Vehicular Embodied AI Networks. *IEEE Transactions on Mobile Computing* 24, 5 (2025), 4573–4588. 3.3
- [262] Yifan Zhong, Jakub Grudzien Kuba, Xidong Feng, Siyi Hu, Jiaming Ji, and Yaodong Yang. 2024. Heterogeneous-Agent Reinforcement Learning. *Journal of Machine Learning Research* 25, 32 (2024), 1–67. 4.2
- [263] Ming Zhou, Jun Luo, Julian Villegas, Yaodong Yang, David Rusu, Jiayu Miao, Weinan Zhang, Montgomery Alban, IMAN FADAKAR, Zheng Chen, Chongxi Huang, Ying Wen, Kimia Hassanzadeh, Daniel Graves, Zhengbang Zhu, Yihan Ni, Nhat Nguyen, Mohamed Elsayed, Haitham Ammar, Alexander Cowen-Rivers, Sanjeevan Ahilan, Zheng Tian, Daniel Palenicek, Kasra Rezaee, Peyman Yadmallat, Kun Shao, dong chen, Baokuan Zhang, Hongbo Zhang, Jianye Hao, Wulong Liu, and Jun Wang. 2021. SMARTS: An Open-Source Scalable

- Multi-Agent RL Training School for Autonomous Driving. In *Proceedings of the 2020 Conference on Robot Learning*. 264–285. 4, 23c, 4.4
- [264] Xuanhan Zhou, Jun Xiong, Haitao Zhao, Xiaoran Liu, Baoquan Ren, Xiaochen Zhang, Jibo Wei, and Hao Yin. 2024. Joint UAV trajectory and communication design with heterogeneous multi-agent reinforcement learning. *Science China Information Sciences* 67, 3 (2024), 132302. 2.4
- [265] Ziyu Zhou, Gang Wang, Jian Sun, Jikai Wang, and Jie Chen. 2023. Efficient and Robust Time-Optimal Trajectory Planning and Control for Agile Quadrotor Flight. *IEEE Robotics and Automation Letters* 8, 12 (2023), 7913–7920. 3.1
- [266] Zhi-Hua Zhou. 2022. Open-environment machine learning. *National Science Review* 9, 8 (2022), nwac123. 5