Generative Multi-Agent Collaboration in Embodied AI: A Systematic Review

Di Wu[1] , **Xian Wei**[3] , **Guang Chen**[1] , **Hao Shen**[4] , **Bo Jin**[1,2*]

[1]Tongji University
[2]Shanghai Research Institute for Intelligent Autonomous Systems
[3]East China Normal University
[4]Technical University of Munich
{wu2002, guangchen, bjin}@tongji.edu.cn, xwei@sei.ecnu.edu.cn, shen@fortiss.org

Abstract

Embodied multi-agent systems (EMAS) have attracted growing attention for their potential to address complex, real-world challenges in areas such as logistics and robotics. Recent advances in foundation models pave the way for generative agents capable of richer communication and adaptive problem-solving. This survey provides a systematic examination of how EMAS can benefit from these generative capabilities. We propose a taxonomy that categorizes EMAS by system architectures and embodiment modalities, emphasizing how collaboration spans both physical and virtual contexts. Central building blocks, perception, planning, communication, and feedback, are then analyzed to illustrate how generative techniques bolster system robustness and flexibility. Through concrete examples, we demonstrate the transformative effects of integrating foundation models into embodied, multi-agent frameworks. Finally, we discuss challenges and future directions, underlining the significant promise of EMAS to reshape the landscape of AI-driven collaboration.

1 Introduction

Embodied multi-agent systems (EMAS) have garnered growing interest due to their significant potential in domains such as smart transportation, logistics, and manufacturing [Yan *et al.*, 2013; Ismail *et al.*, 2018]. By integrating physical embodiments—ranging from autonomous vehicles to robotic manipulators—with multi-agent systems (MAS) [Dorri *et al.*, 2018], EMAS offers a decentralized, collaborative approach that can handle complex tasks with remarkable efficiency. Despite these advantages, designing and implementing effective EMAS remains a non-trivial endeavor, often requiring specialized knowledge of cybernetics, extensive training data, and careful reinforcement-learning paradigms [Landauer and Bellman, 2008; Orr and Dutta, 2023].

In traditional MAS, agents collaborate by dividing responsibilities, sharing state information, and collectively adapting to dynamic environments [Dorri *et al.*, 2018]. While these

---
*Corresponding author

principles have led to impressive success in certain niches, conventional approaches face critical limitations in generalizing to new tasks [Mahajan *et al.*, 2022], scaling to large agent populations [Cui *et al.*, 2022], and coping with unexpected environmental changes [Weinberg and Rosenschein, 2004]. They often rely on narrowly trained models, which can be brittle or constrained to specific domains [Yuan *et al.*, 2023]. These shortfalls underscore the urgency for more flexible and robust solutions that can thrive in open-ended and rapidly changing embodied scenarios.

Recent breakthroughs in foundation models (FMs, e.g., large language models, or vision-language models) [Zhou *et al.*, 2024] have opened new avenues for advancing MAS toward more adaptive and generative behaviors. By equipping agents with natural language capabilities, contextual reasoning, and the ability to generate novel solutions, FM-based MAS transcend some of the limitations inherent in purely signal-driven or reinforcement-based frameworks [Guo *et al.*, 2024a; Chen *et al.*, 2024c; Lu *et al.*, 2024]. These "generative agents" can communicate in semantically rich ways, collaborate with human-level fluency, and rapidly adapt strategies in response to unforeseen challenges. As such, FM-powered agents could transform how multi-agent collaboration unfolds—both in physical spaces populated by embodied devices and in virtual spaces where agents share abstract knowledge and tasks.

Against this backdrop, the field of EMAS stands poised to benefit from these recent advances in FMs. By combining physical embodiments with generative multimodal intelligence, future systems may embrace a broader design space that incorporates complex perception, high-level linguistic and visual reasoning, and adaptive decision-making. However, existing literature surveys on embodied AI and multi-agent systems often treat these fields in isolation, leaving critical gaps at their intersection [Ismail *et al.*, 2018; Duan *et al.*, 2022; Guo *et al.*, 2024a; Ma *et al.*, 2024; Hunt *et al.*, 2024]. A systematic view of how FM-based generative agents can best be integrated into EMAS is still emerging.

This survey aims to provide a comprehensive and structured examination of the current state of generative multi-agent collaboration in Embodied AI, as shown in Figure 1. We introduce a taxonomy that classifies existing EMAS solutions based on the number of models and types of embodi-
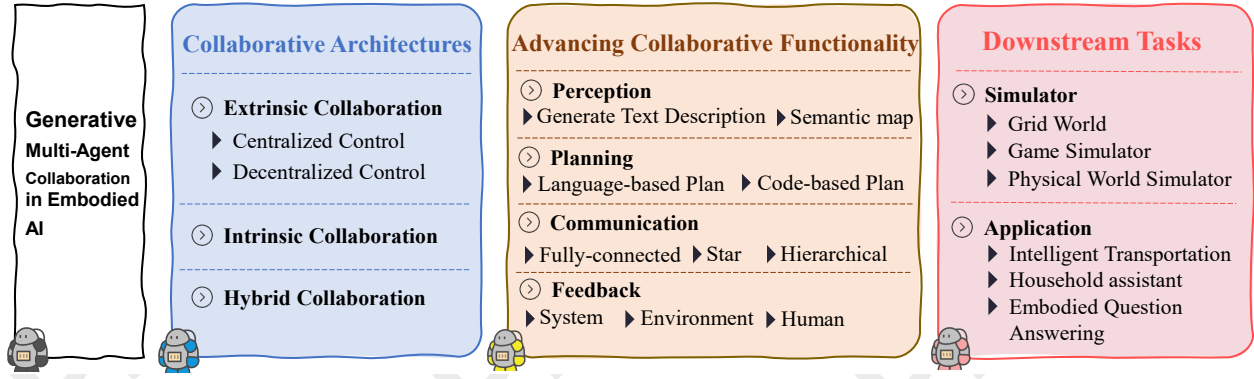
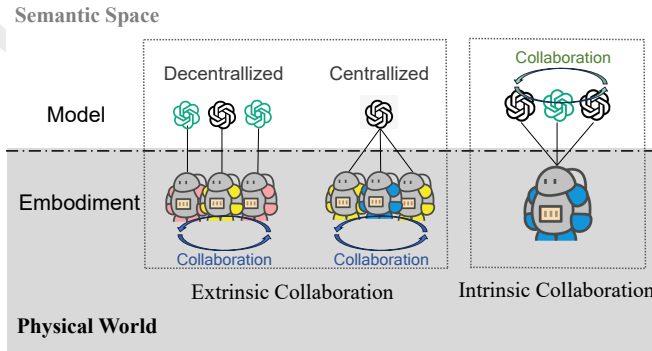Figure 1: A unified multi-agent framework for generative embodied AI.

Figure 2: The embodied multi-agent collaborative architecture.

ments in Section 2, highlighting how collaboration can arise both among physical agents and in purely virtual semantic environments. We explore the major building blocks of multi-agent collaboration—system perception, planning, communication, and feedback—and examine how each of these components can be designed to leverage FM-based generative capabilities in Section 3. Finally, moving beyond theoretical perspectives, we delve into practical applications in Section 4, illustrating how generative multi-agent collaboration enhances functionality across diverse embodied scenarios.

To the best of our knowledge, this is the first survey to systematically address the convergence of MAS, Embodied AI, and foundation models. We close by summarizing open research challenges in Section 5, delineating crucial future directions, and discussing the potential impact of EMAS on broader AI and robotics landscapes. Our goal is to inform and inspire researchers, practitioners, and stakeholders by presenting a holistic overview of this rapidly evolving field.

2 Collaborative Architectures

Building upon the key challenges and opportunities outlined in the previous section, this section presents an overview of collaborative architectures in EMAS, as shown in Figure 2. In particular, we address how generative multi-agent systems leverage either *extrinsic collaboration* across multiple embodied entities or *intrinsic collaboration* among multiple

FMs within a single embodied entity. We also cover *hybrid* approaches that combine these strategies to meet diverse system requirements. The goal is to provide a structured understanding of how multi-agent collaborations can be orchestrated to maximize adaptability, scalability, and task alignment, especially when integrated with FMs.

2.1 Extrinsic Collaboration

In scenarios where collaboration unfolds among multiple embodied entities, known as *extrinsic collaboration*, agents physically or virtually interact to accomplish shared objectives. Drawing from the longstanding multi-robot and conventional MAS literature, extrinsic collaboration can be organized using either a centralized or a decentralized strategy. These approaches offer differing trade-offs related to scalability, communication overhead, and global versus local control.

Centralized Architecture In a centralized policy framework, a single unified model controls multiple robots or agents, offering centralized task allocation and decision-making. The centralized model assigns tasks based on agent capabilities and system objectives, ensuring coordination across agents by providing a global perspective. Studies have explored language-based [Liu *et al.*, 2024b; Obata *et al.*, 2024; Chen *et al.*, 2024b] and code-based [Kannan *et al.*, 2024; Zhang *et al.*, 2024c] task allocation.

The centralized model also plays a key role in decision-making by synthesizing information from all agents to make final decisions, ensuring coherence. For example, [Yu *et al.*, 2023] uses a centralized model for navigation target determination, [Tan *et al.*, 2020] uses it for interactive question answering with a 3D-CNN-LSTM QA model, and [Garg *et al.*, 2024] employs it for deadlock resolution in multi-robot systems by guiding a leader robot's actions.

The centralized control strategy ensures coordination by using a single model for task allocation and decision-making. Its advantages include optimal task distribution and consistent decisions. However, it can be limited by system complexity, high computational demands, and scalability issues in large or dynamic environments.

Decentralized Architecture In a decentralized strategy, each model independently controls its corresponding embodied entity, providing greater flexibility and scalability. Early

Collaboration Architecture		Work	Component					Application	
			Perception	Planning	Communication	Feedback	Agent Role	Environment	Task
Extrinsic Collaboration	Centralized	[Yu et al., 2023]	Semantic map	Language	Star	–	homo-agents	Simulator	Navigation
		[Liu et al., 2024b]	Description	Language	Star	Environment	hetero-agents	Simulator, real-world	Household tasks
		[Chen et al., 2024b]	Semantic map	Language	FC,Star	System	hetero-agents	Simulator	Household tasks
		[Kannan et al., 2024]	Description	Code	Star	-	homo-agents	Simulator, real-world	Household tasks
		[Zhang et al., 2024c]	-	Code (PDDL)	Star	System	hetero-agents	Simulator	Household tasks
		[Tan et al., 2020]	Semantic map	-	Star	-	homo-agents	Simulator	EQA
		[Wang et al., 2024]	Description	Language	Star	Human	homo-agents	Simulator	Household tasks
	Decentralized	[Mandi et al., 2024]	Visual detect	Language	FC	Environment	homo-agents	Simulator, real-world	Household tasks
		[Zhang et al., 2024a]	Semantic map	Language	FC	-	homo-agents	Simulator	Household tasks
		[Zhang et al., 2024b]	World model	Language	-	System	homo-agents	Simulator	Cooperative, Competitive game
		[Agashe et al., 2023]	Descrption	Language	-	System	homo-agents	Game	Overcooked, hanabi game
		[Guo et al., 2024b]	-	Language	Hierarchical	Environment	homo-agents, multi-role llm	Simulator	Household tasks
		[Chen et al., 2024a]	Description	Language	Hierarchical	-	homo-agents	Game	Minecraft creation
		[Chen et al., 2024d]	Description	Language	-	System	homo-agents, multi-role llm	Grid	Multi grid tasks
		[Chen et al., 2023c]	-	Language	Hierarchical	Environment	homo-agents	Game	Minecraft creation
		[Zhang et al., 2024d]	-	Language	FC	System	homo-agents	Simulator,game	Household tasks, overcooked game
		[Tan et al., 2023]	Semantic map	-	-	-	homo-agents	Simulator	EQA
		[Wu et al., 2024]	Description	Language	Hierarchical	Environment	homo-agents	Game	Minecraft navigation
Intrinsic Collaboration		[Shi et al., 2024a]	Semantic Map	Language	-	-	multi-role llm	Simulator	Household tasks
		[Qin et al., 2024]	Description	Language	-	Environment	multi-role llm	Game	Minecraft creation, navigation
		[Nayak et al., 2024]	Description	Language	-	System, Environment	multi-role llm	Simulator	Household tasks
		[Zhang et al., 2023]	Description	-	Hierarchical	System	multi-role llm	Grid	Grid Transportation
		[Liu et al., 2023]	Description	Language	-	System,Human	multi-role llm	Game	Overcooked game
		[Chen et al., 2023b]	Visual detect	Language	-	System	multi-role llm	-	EQA

Table 1: Taxonomy of representative works. "-" denotes that a particular element is not specifically mentioned in this work.

studies used reinforcement learning for decentralized control, but the rise of FMs has enabled agents to handle diverse tasks autonomously [Chen et al., 2024a], forming more advanced decentralized systems.

FMs enhance decentralized systems by leveraging their reasoning capabilities to improve individual decision-making based on local partial observations. For example, [Zhang et al., 2024b] utilizes a world model to assist multi-agent planning, where each individual predicts the behavior of other agents through the world model and infers its own plan. Similarly, [Agashe et al., 2023] introduces an auxiliary theory-of-mind reasoning FM to interpret the actions and needs of partner agents, thereby supporting individual decision-making.

Furthermore, with the reasoning and communication capabilities of FMs, FM-based agents exhibit emergent sociality. [Chen et al., 2023a] reveals that when not explicitly instructed on which strategy to adopt, FM-driven agents primarily follow the average strategy, representing an egalitarian organizational structure among agents. Other research [Guo et al., 2024b; Chen et al., 2024a] highlights the potential benefits of more structured roles within the team. This suggests that, similar to human social structures, FM agents can exhibit emergent behaviors that optimize collaboration by adapting to organizational frameworks, enhancing their collective ability to tackle complex tasks.

2.2 Intrinsic Collaboration

While extrinsic collaboration deals with multiple robots and embodied entities, *intrinsic collaboration* occurs within the internal structure of a single system that may contain multiple FMs. This concept resonates with the recent push for collaborative workflows among various FM modules, each specializing in different roles, to jointly handle increasingly complex tasks. Such internal orchestration expands traditional notions of multi-agent coordination by focusing on consolidated decision-making within a single embodiment.

Each FM in this workflow assumes a specific function or *role* to collaboratively complete a task. Research has applied this paradigm to embodied learning systems, such as [Qin et al., 2024], which uses modules like planner, patroller, and performer for task-solving in a Minecraft sandbox, and [Shi et al., 2024a], which decomposes tasks into observer, planner, and executor roles. LLaMAR [Nayak et al., 2024] also employs a plan-act-correct-verify framework for self-correction without oracles or simulators.

Intrinsic collaboration can improve system functionality by enhancing planning accuracy, safety, and adaptability. For example, [Liu et al., 2023] uses FM-based fast-mind and slow-mind for collaborative plan generation and evaluation, while LLaMAC [Zhang et al., 2023] employs multiple critics and an assessor to provide feedback and improve robustness.

2.3 Hybrid Collaboration Architectures

In many real-world applications, drawing a strict boundary between extrinsic and intrinsic collaboration is neither practical nor advantageous. Instead, hybrid collaborative architectures combine these strategies to exploit the strengths of centralized, decentralized, and internal FM workflows. As embodied tasks grow in complexity, the flexibility to mix different levels of collaboration, both among robots and within an agent's internal structure, becomes increasingly valuable.

Intrinsic collaboration enhances model capabilities through modular FMs and can be applied in centralized and decentralized systems. For example, CoELA [Zhang *et al.*, 2024a] uses five modules—perception, memory, communication, planning, and execution—while [Yu *et al.*, 2024] builds agents with observation, memory, and planning modules for decentralized robot collaboration. Centralized models can also use modular FMs, such as [Wu *et al.*, 2024], which employs a task FM and an action FM for robot task assignment.

Centralized and decentralized strategies can be combined, with different stages of a task utilizing different approaches. Inspired by the centralized training with decentralized execution (CTDE) framework in multi-agent reinforcement learning (MARL), [Chen *et al.*, 2024b] and [Zhao *et al.*, 2024a] propose centralized planning with decentralized execution, where global planning guides task execution, maximizing the synergy between global oversight and local autonomy.

By showcasing these varying architectures, we illustrate how practitioners can effectively orchestrate multi-agent collaboration in EMAS across different levels of granularity and control. The next section builds on this architectural perspective by examining how key system components–perception, planning, communication, and feedback–can be designed to leverage FM-based generative capabilities for more robust and adaptive multi-agent collaboration.

3 Advancing Collaborative Functionality

Building upon the architectural insights from Section 2, which examined how multi-agent collaboration can be orchestrated at the structural level, we now pivot to the functional building blocks that drive effective teamwork among embodied agents. Specifically, we highlight how *perception*, *planning*, *communication*, and *feedback* mechanisms can be designed to harness the generative capabilities of FMs. By focusing on these key modules, we illustrate how EMAS solutions can more robustly interpret the physical environment, formulate and adapt plans, exchange information, and iteratively learn from both their own behaviors and the environment itself. This approach complements the collaboration architectures introduced previously, offering a finer-grained perspective on enabling dynamic and context-aware cooperation among embodied agents.

3.1 Perception

Although a generative model may derive semantic knowledge from text and vision, embodied agents must actively sense and interpret the physical world. This entails handling 3D structures, dynamic conditions, and real-time interactions [Liu *et al.*, 2024e]. Consequently, the perception module is paramount, as it conveys detailed environmental features to subsequent models, ensuring that generative capabilities are grounded in tangible contexts [Pan *et al.*, 2024].

Physical Perception for FM The simplest means of providing physical context to an FM is to supply a verbal description of the environment. Although such prompts may be crafted manually, many approaches augment linguistic descriptions with automated tools. For instance, some studies [Mandi *et al.*, 2024; Chen *et al.*, 2023b] use visual models to detect and describe objects, while others [Brohan *et al.*, 2023; Huang *et al.*, 2023] employ affordance learning to enrich the FM's understanding of how objects can be operated upon in a physical setting. Beyond passively receiving information, recent work enables agents to decide *when* and *what type* of information to observe, facilitating active perception. For example, [Qin *et al.*, 2024] allows the FM to query a fine-tuned model about environmental details; the responses inform a progressively constructed scene description.

Collaborative Perception In multi-agent systems, collaborative perception aims to merge complementary sensory inputs from different agents, enhancing overall performance [Yang *et al.*, 2023]. Within autonomous driving or drone fleets, this often arises through sensor-level data sharing or output-level fusion [Singh *et al.*, 2024]. In FM-based systems, collaborative agents may collectively build a global memory of the environment by aggregating each agent's local maps or visual data. For instance, [Yu *et al.*, 2023] fuses the semantic maps derived from RGBD inputs of multiple agents, and [Tan *et al.*, 2020] employs 3D reconstruction of each agent's observations to form a holistic 3D status and semantic memory of the shared environment.

3.2 Planning

Planning constitutes a core module of multi-agent embodied systems, enabling agents to strategize based on states, goals, and individual capabilities. Effective planning is crucial for task assignment, coordination, and seamlessly integrating the capabilities of generative FMs.

Planning format Planning methods often employ either *language-based* or *code-based* formats. Language-based planning uses natural language to guide task flows, achieving intuitiveness and ease-of-adaptation, especially with the advent of advanced FMs [Mandi *et al.*, 2024; Yu *et al.*, 2023]. By contrast, code-based methods utilize structured programming or domain-specific notations (e.g., PDDL) for higher precision. [Kannan *et al.*, 2024] uses Python code to frame overall task flow, and [Zhang *et al.*, 2024c] converts tasks into PDDL problems for allocation to multiple robots.

Planning process Beyond individual decision-making, multi-agent collaboration demands consensus-building, conflict resolution, and resource sharing. In *centralized* systems, a single model frequently allocates sub-tasks. For example, [Liu *et al.*, 2024b] generates action lists based on each agent's capability, [Obata *et al.*, 2024] integrates FMs and linear programming to solve task partitions, while [Chen *et al.*, 2024b] exploits "robot resumes" for FM-based discussions around task assignment. In *decentralized* systems, agents commu-

nicate directly to optimize their collective plans, supported by robust information exchanges.

3.3 Communication

Communication is central to MAS, enabling agents to share situational updates, coordinate tasks, and reach consensus. Unlike traditional approaches that require painstaking communication protocol design, generative agents can exploit the zero-shot language generation abilities of FMs, reducing the complexity of building efficient communication interfaces.

Following [Sheng *et al.*, 2022], we categorize multi-generative-agent communication patterns in embodied AI to three main structures:

- **Star**: A virtual central agent controls the flow of messages, broadcasting plans or directives to other agents. Much work with centralized architecture has explored this approach [Kannan *et al.*, 2024; Yu *et al.*, 2023]

- **Fully Connected (FC)**: Every agent communicates freely with every other agent, leveraging FM-driven messages. For instance, [Mandi *et al.*, 2024] uses inter-FM dialogues between two robotic arms to coordinate manipulation tasks. In CoELA [Zhang *et al.*, 2024a], each agent references current state information via memory retrieval, generating communication content through an FM.

- **Hierarchical**: A leadership structure emerges to boost scalability and reduce communication overhead. [Chen *et al.*, 2024a; Liu *et al.*, 2024f; Guo *et al.*, 2024b] show how leadership roles channel or filter communications, improving efficiency and outcomes.

3.4 Feedback

Embodied tasks are complex and uncertain, making feedback mechanisms essential for agent improvement. Feedback enables agents to adjust and optimize behavior, allowing continuous learning based on the current state, environmental changes, or external guidance.

System Feedback System feedback refers to information generated internally before an action is taken. This involves agents or a centralized model revisiting their initial plans to identify flaws or potential improvements. Several works [Liu *et al.*, 2024a; Chen *et al.*, 2024b; Zhang *et al.*, 2023] implement a multi-agent discussion phase post-plan generation, refining action lists through peer feedback. [Chen *et al.*, 2024d] and [Zhang *et al.*, 2024c] employ FM checkers to validate code-based plans, ensuring syntactic correctness. Meanwhile, [Zhang *et al.*, 2024d] devises advantage functions to evaluate and iteratively refine plans, and [Liu *et al.*, 2023] applies an FM to predict plan consequences, followed by another FM that rates plan quality, thus driving iterative enhancements.

Environmental Feedback Environmental feedback surfaces after executing actions in the physical (or simulated) world. Many studies log real-world outcomes to guide future decisions. For example, [Liu *et al.*, 2024b] and [Yu *et al.*, 2024] store action results in memory for future planning references, whereas [Qin *et al.*, 2024] and [Nayak *et al.*, 2024] assess the root cause of failures and adapt their action plans

accordingly. Additionally, multi-agent organizational structures can be reconfigured mid-task in response to environmental signals. [Chen *et al.*, 2023c] dynamically updates the roles, and [Guo *et al.*, 2024b] employs a critic FM to evaluate agent performance, even reorganizing leadership.

Human Feedback External human guidance can offer nuanced interventions and strategic directions unattainable through purely automated systems. For instance, [Park *et al.*, 2023] identifies *ambiguous* or *infeasible* task instructions warranting human assistance, while [Wang *et al.*, 2024] and [Ren *et al.*, 2023] integrate conformal prediction to measure task uncertainty and trigger human help requests. Beyond soliciting assistance, [Cui *et al.*, 2023] and [Shi *et al.*, 2024b] permit human operators to refine on-the-fly robot actions through spoken instructions, improving task success rates.

In sum, perception, planning, communication, and feedback emerge as foundational pillars for translating high-level collaborative architectures into practical, generative multi-agent solutions. Whether agents collaborate extrinsically through distributed configurations or intrinsically via multiple FMs within a single embodied, robust supporting modules ensure adaptability and resilience in real-world settings.

The next section delves into concrete application domains, illustrating how these functional modules synergize to tackle diverse embodied tasks. By bridging architectural principles (Section 2) with modular functionalities and grounding them in real-world scenarios, we aim to offer a comprehensive view of how generative multi-agent collaboration can be effectively realized in EMAS.

4 Downstream Tasks: From Simulation to Real-World Deployment

Building on the architectures and functional modules, this section examines how generative multi-agent collaboration moves from controlled simulation environments to real-world applications. Although many advances are validated through virtual platforms, these simulation insights lay the groundwork for tackling the complexities of intelligent transportation, household robotics, and embodied question answering.

4.1 Simulation Platforms

Earlier sections introduced how multi-agent collaboration can be structured and functionally enabled. Simulation environments now enter as a crucial layer for testing these designs, allowing researchers to systematically refine agent interactions without incurring real-world operational costs or risks.

Grid-World Paradigms Grid worlds feature cell-based structures that focus on decision-making and path planning while abstracting away physical details. By adopting an FM-based translator-and-checker framework, [Chen *et al.*, 2024d] improves multi-agent performance on grid tasks, while [Zhang *et al.*, 2023] introduces feedback mechanisms to enhance grid transportation. [Chen *et al.*, 2024e] further evaluates various FM-driven multi-robot architectures in a grid setup, underscoring how these simplified worlds facilitate quick validation of collaborative designs.

Simulator Level	Environment and Benchmark	Related paper
Grid	Grid World	[Zhang *et al.*, 2023], [Chen *et al.*, 2024d], [Chen *et al.*, 2024e]
Game	Minecraft	[Chen *et al.*, 2024a], [Park *et al.*, 2024], [Qin *et al.*, 2024], [Zhao *et al.*, 2024a], [Zhao *et al.*, 2024b]
	Overcooked-AI	[Agashe *et al.*, 2023], [Ying *et al.*, 2024], [Zhang *et al.*, 2024d]
Real-world	ThreeDWorld	[Zhang *et al.*, 2024b], [Liu *et al.*, 2024a], [Zhang *et al.*, 2024a]
	AI-THOR	[Kannan *et al.*, 2024], [Wang *et al.*, 2024], [Nayak *et al.*, 2024], [Zhang *et al.*, 2024c], [Liu *et al.*, 2024d]
	ALFRED	[Shi *et al.*, 2024a]
	Habitat	[Chen *et al.*, 2024b], [Yu *et al.*, 2023]
	VitualHome	[Ying *et al.*, 2024], [Guo *et al.*, 2024b]
	RocoBench	[Mandi *et al.*, 2024], [Zhang *et al.*, 2024d]
	Bestman	[Yu *et al.*, 2024]
	Behavior-1k	[Liu *et al.*, 2024b]

Table 2: Simulators and Applications

Game-Based Collaboration Scenarios Game-based platforms like Overcooked provide clear rules, time constraints, and forced coordination among agents [Ying *et al.*, 2024; Agashe *et al.*, 2023; Zhang *et al.*, 2024d]. FM-coordination extends to other structured games such as Hanabi and Collab Games, showcasing that generative approaches are adaptable to diverse team-based challenges. For more open-ended tasks, Minecraft [Wang *et al.*, 2023; Park *et al.*, 2024] pushes the envelope with larger environments and indefinite goals. Recent work [Park *et al.*, 2024; Zhao *et al.*, 2024a; Qin *et al.*, 2024] focuses on collaborative exploration, while others [Chen *et al.*, 2024a; 2023c; Zhao *et al.*, 2024b] tackle resource collection or structure building.

Advanced 3D Environments and Robotic Simulations Realistic simulators aim to mirror real-life complexity more closely. AI2-THOR [Kolve *et al.*, 2017] offers meticulously rendered indoor scenes and is used for multi-agent household tasks [Kannan *et al.*, 2024; Wang *et al.*, 2024; Liu *et al.*, 2022; Shi *et al.*, 2024a]. Similarly, VirtualHome-Social [Guo *et al.*, 2024b], BEHAVIOR-1K [Liu *et al.*, 2024b], and Habitat-based benchmarks [Chen *et al.*, 2024b] enable agents to develop collaborative strategies for object manipulation and navigation. Such platforms help bridge the gap between algorithmic development and physical deployment.

4.2 Emerging Applications

Armed with validated architectures and robust functional modules, researchers is facing an ultimate frontier: translating simulator learnings into viable physical deployments. From intelligent transportation to household robotics, this subsection spotlights how generative multi-agent collaboration is being adapted to meet real-world demands, illustrating both the maturity and remaining challenges of these systems.

Intelligent Transportation and Delivery Multi-agent collaboration in intelligent transportation covers UAV/UGV coordination for cargo delivery and environmental monitoring. Early approaches mainly leveraged MARL, but FM-driven solutions are now emerging. [Gupte *et al.*, 2024] explores FM-based initial task allocation for surveillance missions, and [Wu *et al.*, 2024] applies generative models to assign tracking targets, suggesting that language-guided strategies can adapt swiftly to dynamic scenarios.

Household Assistance Robotics Many 3D simulation benchmarks, including AI2-THOR and Habitat, were originally crafted to emulate domestic environments. Domestic tasks such as "clearing the table" or following instructions like "Turn on the desk and floor lamp and watch TV" demand robust perception, planning, and communication. Studies [Kannan *et al.*, 2024; Wang *et al.*, 2024; Liu *et al.*, 2024d; Mandi *et al.*, 2024; Zhang *et al.*, 2024d] demonstrate how multiple agents can share roles, interpret commands, and divide complex tasks. Generative models further streamline coordination, enabling adaptive task assignment and richer human-robot interactions.

Beyond Exploration: Embodied Question Answering (EQA) involves active exploration and reasoning in 3D spaces. Unlike tasks that emphasize physical interactions, EQA focuses on gathering and interpreting information, often requiring an advanced understanding of spatial layouts, object relationships, or event histories. Multi-agent versions often leverage team-based sensing for global memory and consensus [Tan *et al.*, 2023; 2020; Patel *et al.*, 2024]. [Chen *et al.*, 2023b] positions agents with specialized functions to contribute key information, showcasing how FM-driven collaboration can integrate observations into coherent answers.

Highlighting these simulation benchmarks and real-world applications, we underscore a key trajectory in EMAS: leveraging structured testbeds for proof-of-concept, then transitioning solutions to high-stakes domains. Having established where and how generative multi-agent collaboration can be deployed, the next sections propose remaining challenges and outline prospective frontiers for EMAS research.

5 Open Challenges and Future Trends

As the field of multi-agent collaboration in embodied AI systems is continuing to develop, there remain several open challenges and promising future directions. Despite the progress made, numerous real-world obstacles persist, limiting the application of embodied systems. This section identifies key challenges and outlines potential areas of exploration and innovation to address these issues.

Benchmarking and Evaluation One major challenge is the lack of standardized evaluation criteria. While significant strides have been made in benchmarking individual agents and single-agent systems, there is a notable gap for the evaluation of embodied multi-agent collaboration. Existing benchmarks focus on task-specific metrics, and fail to account for the complexity of interactions, coordination, and emergent behaviors that arise in multi-agent settings. There is an urgent need for unified evaluation standards for the holistic performance, including factors such as scalability, adaptability, robustness, and collective intelligence. The development of benchmarks is crucial to ensure consistency across different domains, and enabling meaningful comparisons between various multi-agent frameworks.

Data Collection and Heterogeneity Another challenge in multi-agent collaboration is the data scarcity and heterogeneity for embodied systems. Collecting large-scale, high-quality data of different systems with diverse physical characteristics and capabilities is an arduous task. The variation

in hardware, sensor, and environmental interactions leads to inconsistency, making it difficult to generalize across systems and tasks. The real-world data available could be limited, hindering effective training and evaluating. Additionally, most works in multi-agent collaboration are conducted in simulated environments, due to practical constraints, while only a few studies employ real-world data. Hence, there is a pressing need for standardized data collection, as well as innovative methods to transfer between simulation and real-world applications, to bridge the gap between theory and reality.

Foundation Models for Embodied AI The development of foundation models, particularly for embodied agents, is poised to be a transformative breakthrough in the field of multi-agent collaboration. Currently, generative agents primarily rely on FMs to perform complex tasks, and naturally the next step is to build specific foundational models designed for embodied systems. These models serve as a core framework for multi-agent collaboration, integrating perception, decision-making, and action. Recent works, such as RT-1 [Brohan *et al.*, 2022] and RDT [Liu *et al.*, 2024c], made significant strides in robot foundation models for adaptable and scalable systems. The evolution of foundation models will lay the groundwork for more seamless multi-agent collaboration, enabling agents with comprehensive capabilities and teamwork in dynamic environments. However, challenges remain to extend single-agent foundation models to multi-agent, requiring novel architectures and methodologies.

Scalability of Agents Currently, the number of agents involved in collaborative multi-agent systems remains small. Scaling up the number of agents will lead to the increased complexity and difficulty of computation, communication, coordination, task allocations, and resource management. Moreover, maintaining stability and robustness in large-scale multi-agent systems requires sophisticated orchestration and coordination techniques. Research on scalable architecture, efficient communication protocol, and collaborative tactics will be essential to unlocking the full potential of large-scale embodied systems. The development of optimize agent workflows and patterns will be crucial for scaling up these systems in a resource-awareness manner.

Human-Centric Collaboration The integration of robots into human-centered environments remains a critical topic. In many applications, multi-agent systems need to collaborate with not only each other but also human. Ensuring that robots can work seamlessly alongside humans in dynamic and unstructured environments requires the development of human-robot interaction (HRI) protocols that consider human cognitive capabilities, preferences, and limitations. Human-robot collaboration introduces additional challenges, such as safety, adaptability, and trustworthiness. Researches on human-robot teamwork, shared autonomy, and intuitive interfaces will be vital for fostering productive and safe collaboration between humans and robots, particularly for healthcare, industrial automation, and service robots.

Theoretical Foundations and Interpretability Current approaches to embodied multi-agent collaborations, particularly those involving FMs, often lack a solid theoretical foundation. While substantial progress has been made in developing practical systems, the understanding is very limited about the underlying principles and collective intelligence that emerges to govern agent interactions. A deeper theoretical exploration of the dynamic cooperation, including the roles of communication, coordination, and consensus, is essential for advancing the field. Furthermore, the reliability and interpretability of embodied multi-agent systems and models is critical, especially for safety-critical environments, such as autonomous driving and smart railway.

6 Related Work

Although numerous surveys have examined embodied AI or MAS individually, few efforts have tackled the critical overlap between these fields, leaving significant knowledge gaps unaddressed. Early studies on embodied AI [Hu *et al.*, 2023; Firoozi *et al.*, 2023; Ma *et al.*, 2024] focus on single-agent perception-action loops. They discuss autonomy and sensorimotor learning in depth, yet devote limited attention to collaborative paradigms. Similarly, [Duan *et al.*, 2022] and [Liu *et al.*, 2024e] explore how agents interact with their environments, but they still assume solitary agents with limited capacity for distributed teamwork in complex tasks.

In contrast, recent surveys on FM-driven multi-agent systems [Guo *et al.*, 2024a; Chen *et al.*, 2024c; Lu *et al.*, 2024] showcase promising results in semantic communication and emergent coordination, especially in virtual environments. However, these contributions remain detached from physical embodiment, where hardware constraints, sensor noise, and kinematic coordination pose significant challenges. Meanwhile, classic robotics surveys [Ismail *et al.*, 2018; Yan *et al.*, 2013] laid the groundwork for cooperative swarm behaviors but lack generative capabilities that facilitate role adaptation or zero-shot planning.

Several specialized reviews provide partial bridges. For instance, [Hunt *et al.*, 2024] advances language-based human-robot interaction, yet overlooks non-linguistic coordination crucial in industrial or warehouse settings. Likewise, [Sun *et al.*, 2024] integrates FMs with multi-agent reinforcement learning (MARL) but treats embodiment mostly as an implementation detail. Consequently, none of these views examine physical grounding, collaborative intelligence, and generative models under a single unified lens.

Our survey addresses this gap by synthesizing insights from three converging axes: (i) embodied AI's physical imperatives, (ii) multi-agent systems' collaborative intelligence, and (iii) generative models' adaptive reasoning. We propose a novel taxonomy that reconciles embodiment multiplicity (physical agents) with model multiplicity (virtual agents). Through case studies in cross-modal perception and emergent communication, we show how FMs overcome conventional multi-agent limitations in real-world embodied contexts. Hence, we identify underscored challenges, such as Sim2Real transfer for generative collectives, bridging the divide between robotics and FM-based coordination. Our work thus establishes conceptual foundations for a new generation of embodied systems, where physical constraints and generative collaboration progress in tandem.

7 Conclusion

This survey investigates a popular and potential research area, i.e. multi-agent collaboration in embodied systems, and focuses on how generative foundation models can be integrated into embodied multi-agent systems. We emphasize how FM-based generative agents facilitate dynamic collaboration and emergent intelligence, and systematically explore multi-agent collaboration architectures from both intrinsic and extrinsic perspectives, focusing on key technologies such as perception, planning, communication, and feedback mechanisms. Various applications, ranging from grid world exploration to household assistance,are studied to demonstrate the potential of FM-based EMAS in addressing complex problems and to discuss the associated challenges and opportunities in this rapidly evolving field. This survey can serve as a valuable lamp for researchers, practitioners, and stakeholders, offering a comprehensive understanding of the current landscape, and inspires more advanced and scalable solutions of dynamic and seamless collaboration for embodied multi-agent AI.

References

[Agashe *et al.*, 2023] Saaket Agashe, Yue Fan, et al. Evaluating multi-agent coordination abilities in large language models. *arXiv*, 2023.

[Brohan *et al.*, 2022] Anthony Brohan, Noah Brown, et al. Rt-1: Robotics transformer for real-world control at scale. *arXiv*, 2022.

[Brohan *et al.*, 2023] Anthony Brohan, Yevgen Chebotar, Chelsea Finn, Karol Hausman, et al. Do as i can, not as i say: Grounding language in robotic affordances. In *CoRL*, 2023.

[Chen *et al.*, 2023a] Huaben Chen, Wenkang Ji, et al. Multi-agent consensus seeking via large language models. *arXiv*, 2023.

[Chen *et al.*, 2023b] Liang Chen, Yichi Zhang, Shuhuai Ren, Haozhe Zhao, et al. Towards end-to-end embodied decision making with multi-modal large language model: Explorations with gpt4-vision and beyond. In *NeurIPS 2023 Foundation Models for Decision Making Workshop*, 2023.

[Chen *et al.*, 2023c] Weize Chen, Yusheng Su, Jingwei Zuo, et al. Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors. In *ICLR*, 2023.

[Chen *et al.*, 2024a] Jiaqi Chen, Yuxian Jiang, et al. S-agents: self-organizing agents in open-ended environment. *arXiv*, 2024.

[Chen *et al.*, 2024b] Junting Chen, Checheng Yu, Xunzhe Zhou, Tianqi Xu, et al. Emos: Embodiment-aware heterogeneous multi-robot operating system with llm agents. *arXiv*, 2024.

[Chen *et al.*, 2024c] Shuaihang Chen, Yuanxing Liu, Wei Han, Weinan Zhang, and Ting Liu. A survey on multi-generative agent system: Recent advances and new frontiers. *arXiv*, 2024.

[Chen *et al.*, 2024d] Yongchao Chen, Jacob Arkin, Charles Dawson, et al. Autotamp: Autoregressive task and motion planning with llms as translators and checkers. In *ICRA*, 2024.

[Chen *et al.*, 2024e] Yongchao Chen, Jacob Arkin, Yang Zhang, et al. Scalable multi-robot collaboration with large language models: Centralized or decentralized systems? In *ICRA*, 2024.

[Cui *et al.*, 2022] Kai Cui, Anam Tahir, Ekinci, et al. A survey on large-population systems and scalable multi-agent reinforcement learning. *arXiv*, 2022.

[Cui *et al.*, 2023] Yuchen Cui, S. Karamcheti, et al. No, to the right: Online language corrections for robotic manipulation via shared autonomy. In *HRI*, 2023.

[Dorri *et al.*, 2018] Ali Dorri, Salil S Kanhere, and Raja Jurdak. Multi-agent systems: A survey. *IEEE Access*, 6:28573–28593, 2018.

[Duan *et al.*, 2022] Jiafei Duan, Samson Yu, Hui Li Tan, Hongyuan Zhu, and Cheston Tan. A survey of embodied ai: From simulators to research tasks. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(2):230–244, 2022.

[Firoozi *et al.*, 2023] Roya Firoozi, J. Tucker, Stephen Tian, et al. Foundation models in robotics: Applications, challenges, and the future. *The International Journal of Robotics Research*, page 02783649241281508, 2023.

[Garg *et al.*, 2024] Kunal Garg, Jacob Arkin, Songyuan Zhang, Nicholas Roy, and Chuchu Fan. Large language models to the rescue: Deadlock resolution in multi-robot systems. *arXiv*, 2024.

[Guo *et al.*, 2024a] T Guo, X Chen, Y Wang, R Chang, et al. Large language model based multi-agents: A survey of progress and challenges. In *IJCAI*, 2024.

[Guo *et al.*, 2024b] Xudong Guo, Kaixuan Huang, Jiale Liu, et al. Embodied llm agents learn to cooperate in organized teams. *arXiv*, 2024.

[Gupte *et al.*, 2024] Arjun Gupte, Ruiqi Wang, et al. Rebel: Rule-based and experience-enhanced learning with llms for initial task allocation in multi-human multi-robot teams. *arXiv*, 2024.

[Hu *et al.*, 2023] Yafei Hu, Quanting Xie, Vidhi Jain, et al. Toward general-purpose robots via foundation models: A survey and meta-analysis. *arXiv*, 2023.

[Huang *et al.*, 2023] Wenlong Huang, Chen Wang, et al. Voxposer: Composable 3d value maps for robotic manipulation with language models. In *CoRL*, 2023.

[Hunt *et al.*, 2024] William Hunt, Sarvapali D Ramchurn, and Mohammad D Soorati. A survey of language-based communication in robotics. *arXiv*, 2024.

[Ismail *et al.*, 2018] Z H. Ismail, N. Sariff, et al. A survey and analysis of cooperative multi-agent robot systems: challenges and directions. *Applications of Mobile Robots*, 5:8–14, 2018.

[Kannan *et al.*, 2024] Shyam Sundar Kannan, Vishnunandan LN Venkatesh, and Byung-Cheol Min. Smart-llm: Smart multi-agent robot task planning using large language models. In *IROS*, 2024.

[Kolve *et al.*, 2017] Eric Kolve, Roozbeh Mottaghi, et al. AI2-THOR: An Interactive 3D Environment for Visual AI. *ArXiv*, 2017.

[Landauer and Bellman, 2008] C. Landauer and K. L Bellman. Computational embodiment: agents as constructed complex systems. In *Human Cognition and Social Agent Technology*, pages 301–322. 2008.

[Liu *et al.*, 2022] Xinzhu Liu, Xinghang Li, Di Guo, Sinan Tan, Huaping Liu, and Fuchun Sun. Embodied multi-agent task planning from ambiguous instruction. In *RSS*, 2022.

[Liu *et al.*, 2023] Jijia Liu, C Yu, , et al. Llm-powered hierarchical language agent for real-time human-ai coordination. *arXiv*, 2023.

[Liu *et al.*, 2024a] Jie Liu, Pan Zhou, Yingjun Du, Ah-Hwee Tan, et al. Capo: Cooperative plan optimization for efficient embodied multi-agent cooperation. *arXiv*, 2024.

[Liu *et al.*, 2024b] Kehui Liu, Zixin Tang, Dong Wang, Zhigang Wang, et al. Coherent: Collaboration of heterogeneous multi-robot system with large language models. *arXiv*, 2024.

[Liu *et al.*, 2024c] Songming Liu, Lingxuan Wu, et al. Rdt-1b: a diffusion foundation model for bimanual manipulation. *arXiv*, 2024.

[Liu *et al.*, 2024d] Xinzhu Liu, Di Guo, Xinyu Zhang, and Huaping Liu. Heterogeneous embodied multi-agent collaboration. *IEEE Robotics and Automation Letters*, 2024.

[Liu *et al.*, 2024e] Yang Liu, Weixing Chen, Yongjie Bai, et al. Aligning cyber space with physical world: A comprehensive survey on embodied ai. *arXiv*, 2024.

[Liu *et al.*, 2024f] Zhiwei Liu, Weiran Yao, Jianguo Zhang, Liangwei Yang, et al. Agentlite: A lightweight library for building and advancing task-oriented llm agent system. *arXiv*, 2024.

[Lu *et al.*, 2024] Jinliang Lu, Ziliang Pang, et al. Merge, ensemble, and cooperate! a survey on collaborative strategies in the era of large language models. *arXiv*, 2024.

[Ma *et al.*, 2024] Yueen Ma, Zixing Song, et al. A survey on vision-language-action models for embodied ai. *arXiv*, 2024.

[Mahajan *et al.*, 2022] Anuj Mahajan, Mikayel Samvelyan, et al. Generalization in cooperative multi-agent systems. *arXiv*, 2022.

[Mandi *et al.*, 2024] Zhao Mandi, Shreeya Jain, et al. Roco: Dialectic multi-robot collaboration with large language models. In *ICRA*, 2024.

[Nayak *et al.*, 2024] Siddharth Nayak, Adelmo Morrison Orozco, et al. Long-horizon planning for multi-agent robots in partially observable environments. *arXiv*, 2024.

[Obata *et al.*, 2024] Kazuma Obata, Tatsuya Aoki, et al. Lip-llm: Integrating linear programming and dependency graph with large language models for multi-robot task planning. *IEEE Robotics and Automation Letters*, 2024.

[Orr and Dutta, 2023] James Orr and Ayan Dutta. Multi-agent deep reinforcement learning for multi-robot applications: A survey. *Sensors*, 23(7):3625, 2023.

[Pan *et al.*, 2024] Haotian Pan, Shibo Huang, et al. Recent advances in robot navigation via large language models: A review. 2024.

[Park *et al.*, 2023] J. Park, Seungwon Lim, et al. Clara: classifying and disambiguating user commands for reliable interactive robotic agents. *IEEE Robotics and Automation Letters*, 2023.

[Park *et al.*, 2024] J. Park, J. Cho, et al. Mr.steve: Instruction-following agents in minecraft with what-where-when memory. *arXiv*, 2024.

[Patel *et al.*, 2024] Bhrij Patel, Vishnu Sashank Dorbala, et al. Embodied question answering via multi-llm systems. *arXiv*, 2024.

[Qin *et al.*, 2024] Yiran Qin, Enshen Zhou, Qichang Liu, Zhenfei Yin, et al. Mp5: A multi-modal open-ended embodied system in minecraft via active perception. In *CVPR*, 2024.

[Ren *et al.*, 2023] Allen Z Ren, Anushri Dixit, Alexandra Bodrova, et al. Robots that ask for help: Uncertainty alignment for large language model planners. In *CoRL*, 2023.

[Sheng *et al.*, 2022] Junjie Sheng, Xiangfeng Wang, Bo Jin, Junchi Yan, Wenhao Li, et al. Learning structured communication for multi-agent reinforcement learning. *Autonomous Agents and Multi-Agent Systems*, 36(2):50, 2022.

[Shi *et al.*, 2024a] Haochen Shi, Zhiyuan Sun, Xingdi Yuan, et al. Opex: A component-wise analysis of llm-centric agents in embodied instruction following. *arXiv*, 2024.

[Shi *et al.*, 2024b] Lucy X. Shi, Zheyuan Hu, et al. Yell at your robot: Improving on-the-fly from language corrections. *arXiv*, 2024.

[Singh *et al.*, 2024] Apoorv Singh, Gaurav Raut, and Alka Choudhary. Multi-agent collaborative perception for robotic fleet: A systematic review. *arXiv*, 2024.

[Sun *et al.*, 2024] Chuanneng Sun, Songjun Huang, and Dario Pompili. Llm-based multi-agent reinforcement learning: Current and future directions. *arXiv*, 2024.

[Tan *et al.*, 2020] Sinan Tan, Weilai Xiang, Huaping Liu, Di Guo, and Fuchun Sun. Multi-agent embodied question answering in interactive environments. In *ECCV*, 2020.

[Tan *et al.*, 2023] Sinan Tan, Mengmeng Ge, Di Guo, Huaping Liu, and Fuchun Sun. Knowledge-based embodied question answering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10):11948–11960, 2023.

[Wang *et al.*, 2023] Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, et al. Voyager: An open-ended embodied agent with large language models. In *NeurIPS*, 2023.

[Wang *et al.*, 2024] Jun Wang, Guocheng He, and Yiannis Kantaros. Safe task planning for language-instructed multi-robot systems using conformal prediction. *arXiv*, 2024.

[Weinberg and Rosenschein, 2004] M. Weinberg and J.S Rosenschein. Best-response multiagent learning in non-stationary environments. In *AAMAS*, 2004.

[Wu *et al.*, 2024] Yuwei Wu, Yuezhan Tao, Peihan Li, et al. Hierarchical llms in-the-loop optimization for real-time multi-robot target tracking under unknown hazards. *arXiv*, 2024.

[Yan *et al.*, 2013] Zhi Yan, N. Jouandeau, et al. A survey and analysis of multi-robot coordination. *International Journal of Advanced Robotic Systems*, 10(12):399, 2013.

[Yang *et al.*, 2023] Kun Yang, Dingkang Yang, Jingyu Zhang, et al. Spatio-temporal domain awareness for multi-agent collaborative perception. In *ICCV*, 2023.

[Ying *et al.*, 2024] Lance Ying, Kunal Jha, et al. Goma: Proactive embodied cooperative communication via goal-oriented mental alignment. *arXiv*, 2024.

[Yu *et al.*, 2023] Bangguo Yu, Hamidreza Kasaei, and Ming Cao. Co-navgpt: Multi-robot cooperative visual semantic navigation using large language models. *arXiv*, 2023.

[Yu *et al.*, 2024] Wenhao Yu, Jie Peng, Yueliang Ying, et al. Mhrc: Closed-loop decentralized multi-heterogeneous robot collaboration with large language models. *arXiv*, 2024.

[Yuan *et al.*, 2023] Lei Yuan, Ziqian Zhang, et al. Robust multi-agent coordination via evolutionary generation of auxiliary adversarial attackers. In *AAAI*, 2023.

[Zhang *et al.*, 2023] Bin Zhang, Hangyu Mao, Jingqing Ruan, et al. Controlling large language model-based agents for large-scale decision-making: An actor-critic approach. *arXiv*, 2023.

[Zhang *et al.*, 2024a] Hongxin Zhang, Weihua Du, Jiaming Shan, et al. Building cooperative embodied agents modularly with large language models. In *ICLR*, 2024.

[Zhang *et al.*, 2024b] Hongxin Zhang, Zeyuan Wang, Qiushi Lyu, et al. Combo: Compositional world models for embodied multi-agent cooperation. *arXiv*, 2024.

[Zhang *et al.*, 2024c] Xiaopan Zhang, Hao Qin, Fuquan Wang, et al. Lamma-p: Generalizable multi-agent long-horizon task allocation and planning with lm-driven pddl planner. *arXiv*, 2024.

[Zhang *et al.*, 2024d] Yang Zhang, Shixin Yang, Chenjia Bai, Fei Wu, Xiu Li, Zhen Wang, and Xuelong Li. Towards efficient llm grounding for embodied multi-agent collaboration. *arXiv*, 2024.

[Zhao *et al.*, 2024a] Zhonghan Zhao, Kewei Chen, et al. Hierarchical auto-organizing system for open-ended multi-agent navigation. In *ICLR 2024 Workshop on LLM Agents*, 2024.

[Zhao *et al.*, 2024b] Zhonghan Zhao, Ke Ma, Wenhao Chai, Xuan Wang, et al. Do we really need a complex agent system? distill embodied agent into a single model. *arXiv*, 2024.

[Zhou *et al.*, 2024] Ce Zhou, Qian Li, Chen Li, et al. A comprehensive survey on pretrained foundation models: A history from bert to chatgpt. *International Journal of Machine Learning and Cybernetics*, pages 1–65, 2024.