

MeCo: Enhancing LLM-Empowered Multi-Robot Collaboration via Similar Task Memoization

Baiqing Wang
Northwestern Polytechnical
University
Xi’an, China
wbq@mail.nwpu.edu.cn

Helei Cui*
Northwestern Polytechnical
University
Xi’an, China
chl@nwpu.edu.cn

Bo Zhang
Northwestern Polytechnical
University
Xi’an, China
bo.zhang@mail.nwpu.edu.cn

Xiaolong Zheng
Beijing University of Posts and
Telecommunications
Beijing, China
zhengxiaolong@bupt.edu.cn

Bin Guo
Northwestern Polytechnical
University
Xi’an, China
guob@nwpu.edu.cn

Zhiwen Yu
Harbin Engineering University &
Northwestern Polytechnical University
Harbin, China
zhiwenyu@nwpu.edu.cn

ABSTRACT

Multi-robot systems have been widely deployed in real-world applications, providing significant improvements in efficiency and reductions in labor costs. However, most existing multi-robot collaboration methods rely on extensive task-specific training, which limits their adaptability to new or diverse scenarios. Recent research leverages the language understanding and reasoning capabilities of large language models (LLMs) to enable more flexible collaboration without specialized training. Yet, current LLM-empowered approaches remain inefficient: when confronted with identical or similar tasks, they must replan from scratch because they omit task-level similarities. To address this limitation, we propose MeCo, a similarity-aware multi-robot collaboration framework that applies the principle of “cache and reuse” (a.k.a., memoization) to reduce redundant computation. Unlike simple task repetition, identifying and reusing solutions for similar but not identical tasks is far more challenging, particularly in multi-robot settings. To this end, MeCo introduces a new similarity testing method that retrieves previously solved tasks with high relevance, enabling effective plan reuse without re-invoking LLMs. Furthermore, we present MeCoBench, the first benchmark designed to evaluate performance on similar-task collaboration scenarios. Experimental results show that MeCo substantially reduces planning costs and improves success rates compared with state-of-the-art approaches.

KEYWORDS

Multi-robot collaboration; Large language models; Similar tasks

ACM Reference Format:

Baiqing Wang, Helei Cui, Bo Zhang, Xiaolong Zheng, Bin Guo, and Zhiwen Yu. 2026. MeCo: Enhancing LLM-Empowered Multi-Robot Collaboration via Similar Task Memoization. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 9 pages. <https://doi.org/10.65109/QPUC1624>

*Corresponding author.



This work is licensed under a Creative Commons Attribution International 4.0 License.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). <https://doi.org/10.65109/QPUC1624>

1 INTRODUCTION

Multi-robot systems have appeared in many real-world scenarios, like automatic manufacturing [13, 18, 38], smart housekeeping [20, 24, 26], and intelligent healthcare [4, 19, 32]. They significantly improve production efficiency and reduce labor costs. Current methods, including teach programming [27], reinforcement learning (RL) [17, 44], and imitation learning (IL) [14], rely on fixed programming or prolonged training to generate collision-free trajectories for specific tasks, limiting their adaptability across diverse tasks. By leveraging the superior text comprehension, reasoning, and generation capabilities of large language models (LLMs) [6, 9, 30, 31], multi-robot systems achieve flexible handling of diverse coordination scenarios without any task-specific training.

In this direction, a number of LLM-empowered multi-robot collaboration frameworks are proposed, e.g., RoCo [25], EMOS [8], and ReCA [36]. They assign robots to LLM agents, which plan tasks either in a centralized manner or via dialogue-based coordination. With the ability to capture common-sense knowledge [46], LLMs empower agents to provide robots with intelligent task-decomposition and flexible decision-making. Despite promising, the above-mentioned studies all treat tasks as isolated ones and overlook the inter-task relationships.

Similar tasks often appear in real practice. Examples include parcel sorting [1] and product categorization [2] in assembly lines, as well as floor sweeping, sandwich making, and grocery packing in home automation [26]. While task-specific elements (e.g., target object types and positions) may vary to some extent, the core objectives (e.g., sorting parcels into containers or sweeping items into baskets) and inherent constraints (e.g., robot quantity and fixed resources like containers) remain unchanged. However, existing LLM-empowered methods still re-invoke LLM agents when encountering similar tasks, leading to unnecessary costs, i.e., token consumption and planning time. Although some approaches consider reducing tokens, such as HMAS-2 [11], S-ATLAS [37], and ReAd [45], they can hardly cut the planning costs caused by redundant replanning for similar tasks.

In light of these, we propose MeCo, a similarity-aware multi-robot collaboration framework that leverages task-level similarity for efficient planning. However, compared with dealing with exactly the same tasks, finding and leveraging the similarity between two

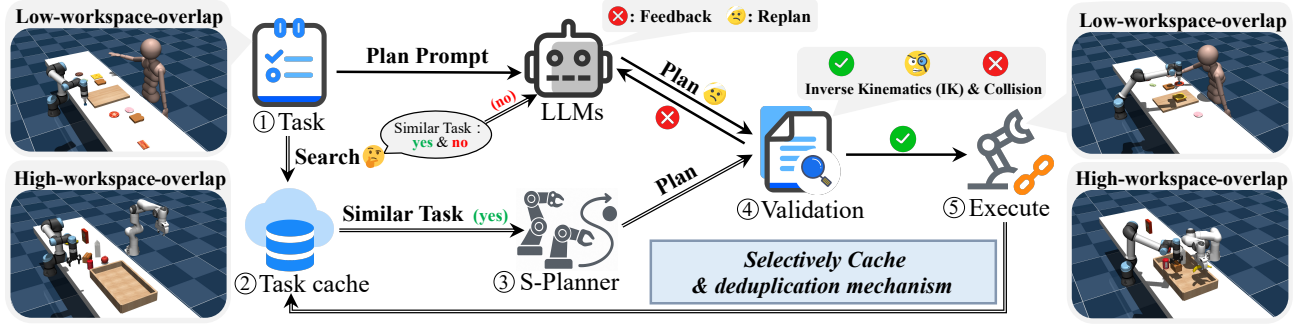


Figure 1: Workflow of MeCo. The arrow \rightarrow shows the LLM-empowered planning; the arrow \Rightarrow shows the MeCo extensions.

tasks is more challenging. In response, a new similarity testing method is designed to identify previously solved tasks with high relevance, enabling the reuse of existing plans without re-invoking the LLMs. As shown in Figure 1, upon receiving a task, MeCo first searches the task cache for a similar task. If no similar task is found, it invokes LLMs for planning, just like the previous work. If an identical or similar task is matched, we further involve a similar motion planner (S-Planner), which generates a plan for the current task by referencing the similar task without relying on LLMs. If S-Planner fails, the failure reason is fed back to the LLMs, which continue planning from the failed step instead of restarting the entire process. Once planning is successful, MeCo selectively stores this task in the task cache. To facilitate evaluation, we further introduce a variant of RoCoBench [25] for similar tasks, called MeCoBench. Experiments on MeCoBench show that our proposed design achieves an overall improvement of around 30% in success rate, about 55% saving in planning time, and up to 70% reduction in token consumption, compared with the state-of-the-art design.

Contribution. To the best of our knowledge, we are the first to explore the problem of similar tasks in multi-robot collaboration. The primary contributions of our work are outlined as follows:

- We propose MeCo, a similarity-aware multi-robot collaboration framework that leverages task-level similarity to enhance task planning powered by LLMs.
- We present a similarity testing method that employs multiple strategies to accurately evaluate task similarity, based on the degree of workspace overlap among robots.
- We design S-Planner, a similar motion planner that references previously executed similar tasks to plan new ones, thereby avoiding repeated LLM invocations.
- We introduce MeCoBench, an extended benchmark designed to evaluate performance on similar tasks. Experimental results demonstrate that MeCo significantly reduces planning costs while improving task success rates.

2 RELATED WORK

LLMs for Robotics. Early prior work uses LLMs to select skill primitives and complete robotic tasks, such as SayCan [5] and Inner Monologue [15]. Later work utilizes LLMs to generate code-formatted robotic policies, such as ProgGPT [34], Code-As-Policy [21], and ELLMER [28]. These works primarily focus on single-robot

setups. Recent advancements extend LLM applications to multi-robot scenarios. For instance, Zhang *et al.* [43] utilize LLMs as planners to facilitate agent collaboration in dialogue-driven tasks. Zhao *et al.* [25] propose an LLM-based multi-robot framework (RoCo), assigning each robot to an LLM agent for task collaboration in a dialogue style. Wan *et al.* [36] introduce a characterization and co-design framework (ReCA), which improves task efficiency and system scalability in LLM-based multi-robot collaboration, while also reducing token usage. Wang *et al.* [37] develop an LLM-based distributed multi-robot planner using conformal prediction, which achieves the user-specified success rate while minimizing token usage. Chen *et al.* [11] compare four LLM planning frameworks in multi-robot coordination to achieve higher token efficiency. While prior work achieves effective multi-robot coordination via LLMs and proposes token cost and planning time reduction methods, scenarios containing similar tasks still require repeated LLMs planning, resulting in unnecessary token costs and planning time.

Learning-based methods for Robotics. Learning-based methods, such as reinforcement learning (RL) [29, 40] and imitation learning (IL) [14, 33], have been applied to multi-robot collaboration. These methods typically require task-specific training to allocate actions and plan paths. For example, Xie *et al.* [41] propose a deep imitation learning framework for robotic bimanual manipulation in a continuous state-action space. Cui *et al.* [12] introduce an actor-critic-based deep reinforcement learning framework for closed-chain manipulation in dual-arm robotic systems. Although learning-based methods have achieved success, in multi-robot collaboration, they require modifications or retraining when tasks change, making them unsuitable for scenarios with similar tasks.

Programming methods for Robotics. Programming methods for robotics are widely used in intelligent manufacturing, such as part assembly on industrial production lines and fixed-item sorting. These methods typically adopt two paradigms: offline programming [3] for predefined trajectories and teach pendant programming [27] for manual path recording. However, their reliance on static task assumptions limits applicability in environments with evolving operational demands.

3 PROBLEM STATEMENT

3.1 LLM-Empowered Planning

We use LLMs to generate 3D waypoint paths that guide low-level motion planning. This practice is common, such as in RoCo [25],

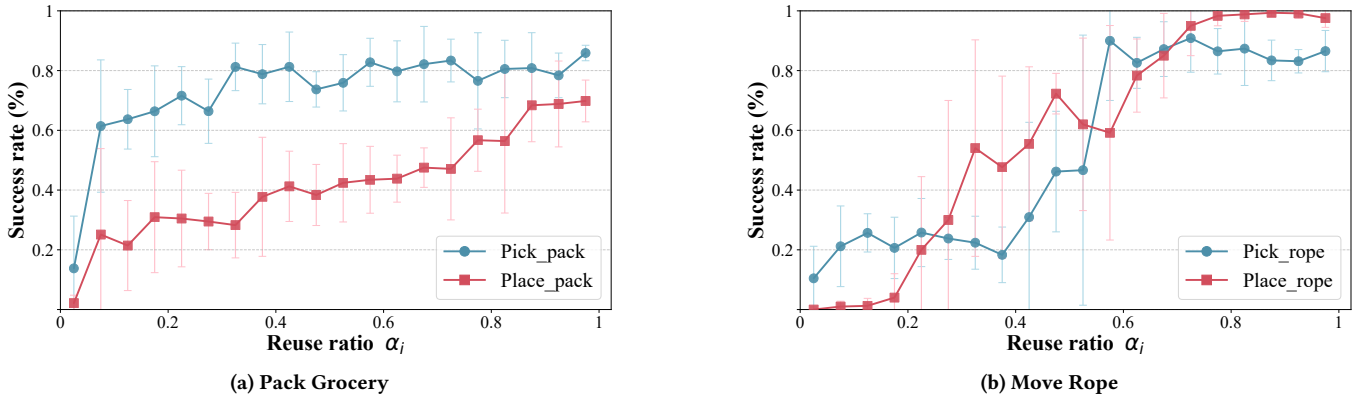


Figure 2: The value of α_i affects the success rate of planning the current task based on similar tasks. We test the “Pick” and “Place” subtasks of the *Pack Grocery* and *Move Rope* tasks under different intervals, with each setting tested 100 times on average.

ReAd [45], AO-Planner [7], and [16, 22]. As shown in the single-line section of Figure 1, the task description is provided to the LLMs in text form. LLMs decompose the task into multiple subtasks. For each subtask, LLMs output 3D waypoint paths in the robots’ workspace to guide the low-level motion planner. The generated 3D waypoints undergo inverse kinematics (IK) and collision checking to ensure their validity. If all waypoints pass validation, the low-level motion planner is invoked to plan and execute the trajectory. Otherwise, the failure reason is fed back to the LLMs for replanning.

3.2 Definition of Tasks

A task is defined as a tuple $T = (C, E)$. The task content C is a text-based specification of execution requirements. The environmental elements E define the task space Ω through $E = (R, O, D, K)$, where $R = \{\{r_i = (\tau_i, n_i, W_i)\}\}_{i=1}^N$ represents robots, with each r_i comprising type τ_i , quantity n_i , and workspace W_i ($W_i \in \Omega$); $O = \{o_j = (p_j, \gamma_j)\}_{j=1}^M$ denotes target objects, where each o_j includes position p_j and category γ_j ; D captures obstacle distributions; and S encodes task-specific constraints, such as recipes or ordering constraints.

For demonstration purposes, we focus on the six tasks in RoCoBench [25], a state-of-the-art benchmark of multi-robot collaboration scenarios, which is widely used in previous studies [10, 22, 45]. RoCoBench consists of six multi-robot collaboration tasks in a tabletop manipulation environment, including *Move Rope*, *Pack Grocery*, *Make Sandwich*, *Sort Cubes*, *Arrange Cabinet*, and *Sweep Floor*. Each task imposes specific workspace constraints on the robots involved. We categorize tasks into two groups based on robots’ workspace overlap levels: **low-workspace-overlap tasks** and **high-workspace-overlap tasks**. Tasks such as *Sweep Floor*, *Make Sandwich*, *Sort Cubes*, and *Arrange Cabinet* fall under low-workspace-overlap tasks, while *Pack Grocery* and *Move Rope* belong to high-workspace-overlap tasks.

4 OUR PROPOSED DESIGN

4.1 Overview

As shown in Figure 1, after each successful planning, MeCo selectively stores the task plan in the task cache for future reference. When planning a new task, it first searches the task cache for a

similar task that meets the current requirements. If no similar task is found, MeCo calls LLMs and uses the task description as the prompt for planning. If a similar task is matched, it then invokes the similar motion planner, namely S-Planner, which generates a plan for the current task efficiently by referencing the similar task without relying on LLMs. S-Planner or LLMs decompose the task into multiple subtasks. For each subtask, they output 3D waypoint paths to guide the low-level motion planner. The generated plan undergoes a set of validations, such as inverse kinematics (IK) and collision checking. If validation fails, the failure reason is fed back to LLMs for replanning. Specifically, if S-Planner fails to generate a valid plan, LLMs continue planning from the failed step instead of restarting from scratch. Once all validations are passed, the final motion trajectories for all robots are planned by the low-level motion planner and executed in the environment.

4.2 Definition of Similar Tasks

In the search phase of MeCo, it needs to identify whether there are any tasks in the task cache that are similar to the current task and select the most similar task for use in subsequent stages. We have defined task similarity criteria separately for high-workspace-overlap tasks and low-workspace-overlap tasks.

For low-workspace-overlap tasks, the task space Ω is partitioned into a finite set of disjoint work areas $\mathcal{A} = \{A_1, \dots, A_M\}$, such as the left and right regions of cutting boards in *Make Sandwich* or the regions between panels in *Sort Cubes*. The historical task is denoted as $T_{\text{ref}} = (C, E_{\text{ref}})$, and the current task as $T_{\text{cur}} = (C, E_{\text{cur}})$. Both tasks share the same C, R, D , and S . T_{cur} and T_{ref} are considered similar if there exists a point mapping, as shown in Eq. (1), that ensures each target object undergoes only restricted geometric transformations within the same region, without altering the cross-region interaction structure.

$$\Pi : \Omega \rightarrow \Omega, \text{ s.t. } \forall j, p_j^{\text{cur}} = \Pi(p_j^{\text{ref}}), p_j^{\text{ref}} \in A_m \Rightarrow p_j^{\text{cur}} \in A_m, \quad (1)$$

where p_j^{ref} and p_j^{cur} denote the position of the j -th target object in the historical and current tasks, respectively.

For high-workspace-overlap tasks, there is a higher risk of collisions. Therefore, it is necessary to further consider the reusability of historical trajectories in the current task environment. The

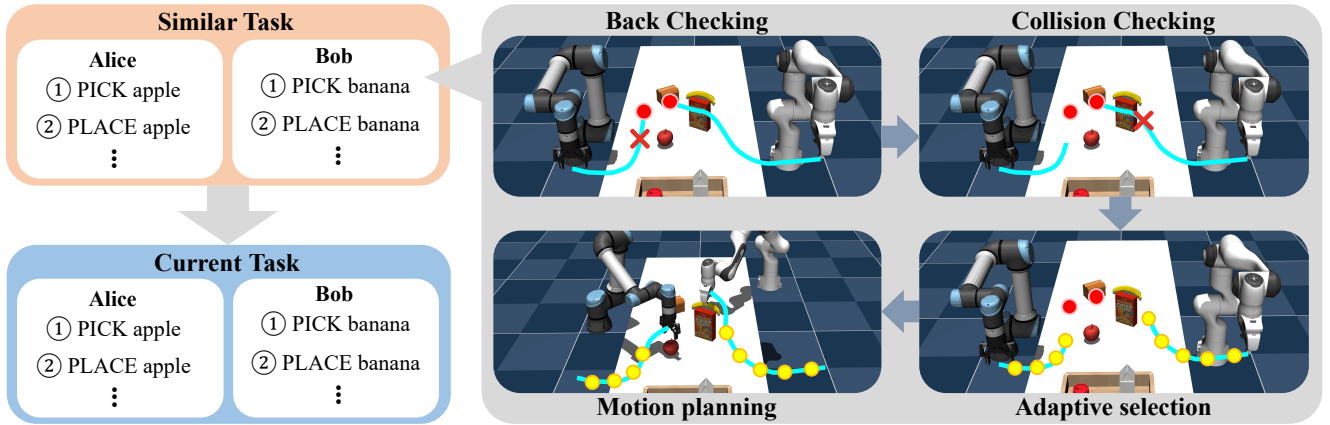


Figure 3: Workflow of S-Planner in the high-workspace-overlap mode. The left side illustrates how tasks are decomposed and assigned at the high level, while the right side shows how motion planning is conducted at the low level. The blue curves represent the reference trajectory, the red points mark the target positions in the similar task; ‘Alice’ and ‘Bob’ refer to robots.

historical task $T_{\text{ref}} = (C, E_{\text{ref}})$ and the current task $T_{\text{cur}} = (C, E_{\text{cur}})$ share the same C, R, D , and S . A task is decomposed into several subtasks, each associated with a reference trajectory of the robot end-effector on the horizontal plane. For the i -th subtask, we define:

- Δx_i^{goal} : the horizontal distance from the target object to its goal position in the current task;
- Δx_i^{ref} : the horizontal length of the reference trajectory from the historical task that is reusable in the current environment.

Accordingly, the reuse ratio α_i is defined as:

$$\alpha_i = \frac{\Delta x_i^{\text{ref}}}{\Delta x_i^{\text{goal}}}, \quad (2)$$

Furthermore, the overall task similarity metric is defined as the average reuse ratio across all subtasks:

$$\alpha = \frac{1}{n} \sum_{i=1}^n \alpha_i, \quad (3)$$

where n denotes the number of subtasks in the historical task plan.

Let p denote the unit failure rate of reusing historical trajectories in the current environment. If the reusable trajectory length is Δx_i^{ref} , its success rate is $(1-p)^{\Delta x_i^{\text{ref}}}$. For the remaining trajectory length $\Delta x_i^{\text{goal}} - \Delta x_i^{\text{ref}}$, a planner generates new trajectories. Let \tilde{p} denote the unit failure rate of replanning. Since replanning is more difficult, $\tilde{p} > p$. The success rate of the replanned segment is therefore $(1-\tilde{p})^{\Delta x_i^{\text{goal}} - \Delta x_i^{\text{ref}}}$. By multiplying the two terms and substituting $\Delta x_i^{\text{final}} = \alpha_i \Delta x_i^{\text{goal}}$, we obtain the lower bound of the subtask success probability:

$$P_{\text{succ},i} \geq (1-\tilde{p})^{\Delta x_i^{\text{goal}}} \cdot \exp(\kappa \cdot \alpha_i \cdot \Delta x_i^{\text{goal}}), \quad (4)$$

where $\kappa = \ln \frac{1-p}{1-\tilde{p}} > 0$. This inequality shows that $P_{\text{succ},i}$ increases monotonically with α_i . Furthermore, we conducted experiments on two sub-tasks, "Pick" and "Place", within the high-workspace-overlap tasks of *Pack Grocery* and *Move Rope*, as shown in Figure 2. The results indicate that as α_i increases, the success rate of sub-tasks

improves. Therefore, we use α as the indicator for determining the similarity of high-workspace-overlap tasks.

MeCo can select an appropriate threshold τ based on its requirements. If $\alpha < \tau$, the tasks are considered dissimilar; otherwise, they are considered similar. For lower values of τ , MeCo relaxes the similarity criteria for task selection, resulting in a lower success rate. In contrast, a higher τ increases the stringency of task selection, leading to a higher success rate. We found that for the *Pack Grocery*, $\alpha = 0.85$, and for the *Move Rope*, $\alpha = 0.7$, the success rate shows little fluctuation while maintaining a high level of performance. Therefore, we choose 0.85 and 0.7 as the thresholds for these two tasks in our subsequent experiments.

4.3 Similar Motion Planning

We now introduce the S-Planner for similar motion planning. Briefly, S-Planner plans the current task by referring to the plans of similar tasks. It operates in two modes: one for tasks with high workspace overlap and the other for tasks with low workspace overlap.

Mode for low-workspace-overlap tasks. By referring to the similar task T_{ref} , S-Planner decomposes the current task T_{cur} into multiple subtasks. Each subtask corresponds to an action sequence $\{a_1, a_2, \dots, a_n\}$, where n denotes the number of robots. At the high level, S-Planner inherits the action sequence from the similar task and applies it to the current task: $a_i^{\text{cur}} = a_i^{\text{ref}}, \forall i \in \{1, \dots, n\}$. At the low level of motion planning, we define the target pose difference based on the target pose g_s :

$$\Delta(g_s^{\text{ref}}, g_s^{\text{cur}}) = \|g_s^{\text{ref}} - g_s^{\text{cur}}\|, \quad (5)$$

where $\|\cdot\|$ is a pose difference metric that combines translational and rotational distances. We then define the decision function as

$$\delta_s = 1 \mathbb{1}[\Delta(g_s^{\text{ref}}, g_s^{\text{cur}}) \leq \varepsilon], \quad (6)$$

where $\varepsilon > 0$ is the tolerance threshold.

- If $\delta_s = 1$: the target pose remains unchanged, and the trajectory is reused directly.
- If $\delta_s = 0$: the target pose changes, and the RRT-based motion planner is invoked to generate a new trajectory.

Mode for high-workspace-overlap tasks. At a high level, S-Planner follows the same approach as the mode for low-workspace-overlap tasks, inheriting the action sequence from the similar task and applies it to the current task: $a_i^{\text{cur}} = a_i^{\text{ref}}, \forall i \in \{1, \dots, n\}$. However, since high-workspace-overlap tasks have a higher risk of collision, the low-level planning method used for low-workspace-overlap tasks is not applicable. In particular, the motion trajectory in the similar task plan may not always be applicable to the current task. For instance, the positions of non-target objects in the current task may change and overlap with the reference trajectory. Therefore, it is essential to identify the valid reference segments of the similar task trajectory, which involves two steps: back checking and collision checking, as shown in Figure 3.

Let the joint-space sequence generated by the historical plan be $Q = \{q_s\}_{s=0}^N$, where each q_s is a set of joint angles. Using forward kinematics, $\text{FK} : \mathbb{R}^{dq} \rightarrow \mathbb{R}^3$, the corresponding end-effector trajectory is $R = \{P_s = (x_s, y_s, z_s)\}_{s=0}^N$. The first step, back checking, prevents unnecessary backtracking during task execution. Specifically, if the target position P_c of the current task is located in the middle of the reference path R , the path is truncated to obtain R . This ensures that the trajectory does not return to previously visited regions. Accordingly, the truncated path is defined as follows:

$$R_b = (P'_s = (x_s, y_s, z_s))_{s: d(x_s - x_c) \geq 0}, \quad (7)$$

where $R_b \subseteq R$, and $d := \text{sign}(x_N - x_0) \in \{+1, -1\}$ denoting the monotonic direction indicator. Next, collision checking is performed on R_b to verify whether the reference motion trajectory is collision-free in the current task environment. If collisions are detected, R_b is truncated before the collision point to obtain R_{final} , where $R_{\text{final}} \subseteq R_b$, ensuring a feasible reference trajectory.

During trajectory planning, S-Planner adaptively selects key sampling points from R_{final} to reduce redundancy while preserving critical information. The highest point along the z-axis, P_{peak} , is identified as the central key point. Using an expansion coefficient η , the expansion step size is computed as $\Delta = \lfloor \eta \cdot N \rfloor$, where N is the total number of points in the path. Based on Δ , expansion continues in both directions within R_{final} until reaching the endpoints on both sides, forming the point sets K_{left} and K_{right} . Thus, the final set of key sampling points is defined as $K = \{P_{\text{peak}}\} \cup K_{\text{left}} \cup K_{\text{right}}$. Finally, K is provided as an important reference to the RRT-based motion planner for generating motion trajectories for all robots.

4.4 Continuous Planning

Although S-Planner can plan by referencing previous task plans without calling LLMs, its effectiveness is influenced by the similarity between the previous and current tasks. If S-Planner fails to generate a valid plan, MeCo reassigns the task to LLMs for replanning. Since S-Planner failures typically occur at intermediate steps, having LLMs replan from scratch would render the previous S-Planner planning efforts ineffective. To address it, we design a continuous planning module. When S-Planner fails, the module stores the current environment locally. Upon invoking LLMs for replanning, it first loads the stored environment and incorporates it into the prompt, ensuring that LLMs continue planning from the failure step rather than starting over. This allows MeCo to

seamlessly switch between two planning modules, significantly enhancing the collaborative efficiency between S-Planner and LLMs.

4.5 Selective Caching

The number of tasks in the task cache affects the performance of the search process. As the number of tasks in the task cache increases, the time complexity of task matching also rises. Therefore, we set a maximum task count, denoted as k . When the task count reaches k , tasks in the cache that are too similar or redundant will reduce the diversity of the tasks, preventing MeCo from efficiently leveraging the similarity between tasks for planning in diverse scenarios. In this case, a deduplication mechanism becomes crucial. Inspired by the least frequently used (LFU) policy [39], we define $f(T_i)$ as the frequency with which task T_i is used as a similar task. The task cache contains two tables to store task frequencies: one is a cache table for speeding up access, and the other is a global table for frequency updates. When planning a new task, MeCo first searches the cache table. If a similar task is found, $f(T_i)$ is incremented by 1, and the updated value of $f(T_i)$ is synchronized to the global table. If no similar task is found, the system invokes the LLMs for planning while asynchronously searching the global table and recording results for future deduplication. After completing a task, MeCo updates both the cache table and the global table based on task usage frequency, using the following principle:

$$C'_{\text{cache}} = \arg \max_{C \subseteq C_{\text{global}}} \sum_{T \in C} f(T), \quad \text{s.t. } |C| = k \quad (8)$$

where C'_{cache} denotes the updated cache table, C_{global} represents the global table, and k is the maximum task count in the cache.

The quality of tasks in the task cache also affects the performance of S-Planner. For example, in *Pack Grocery* task, objects need to be placed inside a box. If an object is initially placed outside the box and then moved inside in a later step, the task is still considered completed. However, when such a task is referenced as a similar task, it introduces redundant steps and may even lead to planning failures. To address this issue, after a task is completed, we first check its effective step count S . If $S \leq S_{\text{base}}$, it passes the check, where S_{base} is the baseline step count.

5 EXPERIMENTAL EVALUATION

We design a series of experiments to validate our methods. In Section 5.1, we first introduce the experimental setup. In Section 5.2, we analyze the impact of task counts stored in the cache on MeCo’s performance and determine the optimal threshold for the maximum number of tasks in the cache. In Section 5.3, we evaluate MeCo’s performance by comparing with four strong baselines. In Section 5.4, we conduct ablation studies on different components of MeCo, demonstrating the effectiveness of each design.

5.1 Experimental Setup

Experimental Environment. We evaluate the performance of our methods on a server with a 24-core Intel Core i9-14900HX CPU, and we use DeepSeek-V3 [23] as the basic LLM policy for all experiments. The simulation is based on MuJoCo [35] physics engine.

MeCoBench. We present a variant of RoCoBench [25] for similar tasks, called MeCoBench. In contrast to RoCoBench, which

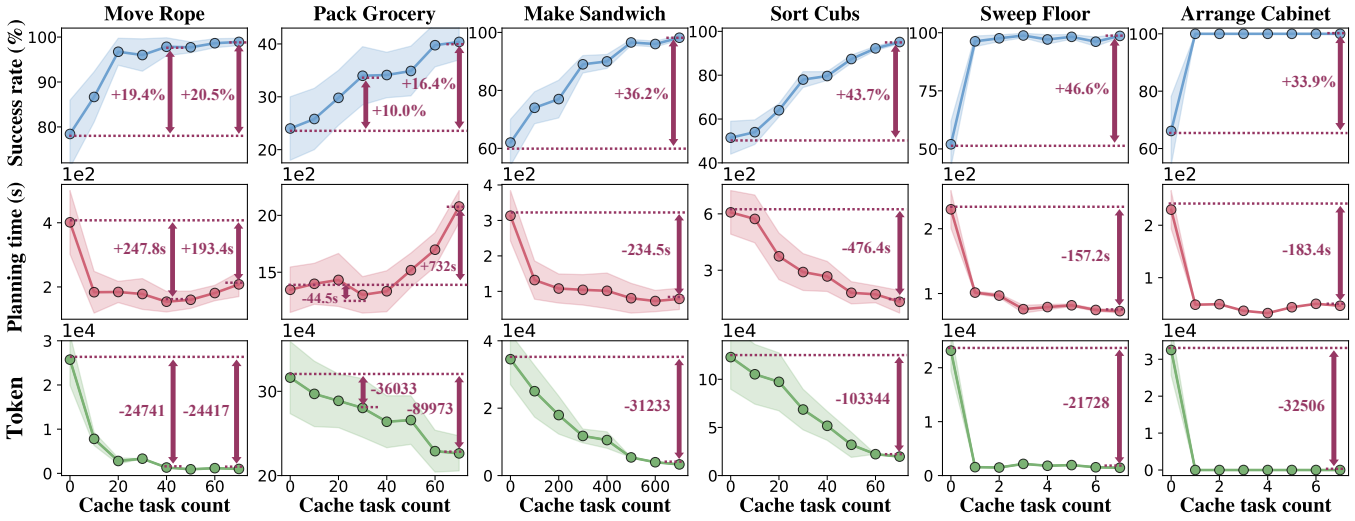


Figure 4: Performance vs. cache task count. As the number of tasks stored in the cache increases, we observe changes in the average success rate, planning time, and token consumption of MeCo across different tasks. For each task type, we randomly store tasks planned by LLM-empowered methods in the task cache and evaluate MeCo averaged over 30 random seeds.

primarily focuses on each task in isolation, MeCoBench emphasizes inter-task similarity. For each task, we introduce similarity indicators to increase benchmark complexity. This allows MeCoBench to operate in three modes: randomly generating tasks that are similar to those in the task cache, different to them, or without considering similarity at all. More details can be found in the Appendix A¹.

Baseline Methods. To evaluate the effectiveness of our framework, we compare it with the following baseline methods:

- RoCo [25]: The baseline restricts each LLM agent to partial information and performs planning through dialogue. It achieves state-of-the-art performance in RoCoBench [25].
- Central Plan: The baseline allows the LLM agent to access full environment observations and knowledge of all robots. It generates actions for all robots at once.
- HMAS-2 [11]: The baseline first lets the central LLM planner generate a set of actions for each robot. Each robot is also equipped with an LLM agent that checks the assigned actions and provides feedback to the central planner, which then replans accordingly.
- ReAct [42]: The baseline enables the LLM agent to generate reasoning traces and task-specific actions in an interleaved manner.

Evaluation Metric. In a finite number of rounds with a limited number of attempts per round, we evaluate performance using the following metrics: (i) **Success rate**: the success rate of task planning, which assesses the algorithm’s planning effectiveness. (ii) **Planning time**: the average time required from the start to the completion of planning, which measures planning efficiency. (iii) **Tokens**: The average token consumption, which reflects the algorithm’s token usage. An algorithm performs better if it achieves a higher Success rate, lower Planning time, and lower Tokens. In addition, we use the standard deviation to measure the variability of success rates. However, once the cache reaches a certain number of tasks, the use of LLMs decreases. This leads to many small values and a few

large values for **Planning time** and **Tokens**. In such cases, the standard deviation can even exceed the mean, making it unsuitable. To address this, we define a truncated standard deviation, which ignores outliers that deviate from the mean by more than τ times the standard deviation. This reduces the impact of extreme values on variability. In our experiments, we set $\tau = 5$.

5.2 Performance vs. Cache Task Count

As the number of stored tasks in the task cache increases, task diversity also improves, enabling more efficient planning through task-level similarity in a wider range of scenarios. However, this also increases the time complexity during the search phase, leading to longer task-matching times. In this experiment, we study the impact of cache task count on MeCo’s performance, as shown in Figure 4. Meanwhile, to ensure higher success rates, shorter planning times, and lower token consumption, it is necessary to choose an appropriate maximum cache size for each task.

High-workspace-overlap tasks include *Move Rope* and *Pack Grocery*. As shown in the first two columns of Figure 4, increasing the cache task count significantly improves success rates while substantially reducing token consumption. This indicates that a larger cache allows the system to better exploit task-level similarity across diverse scenarios, thereby reducing reliance on LLMs and improving planning efficiency. However, planning time follows a “decrease-then-increase” trend. With a small cache task count, the acceleration effect dominates and planning time drops significantly. As the cache continues to grow, the enlarged search space increases the cost of matching similar tasks, which gradually offsets and eventually surpasses the benefits of caching. For example, in *Pack Grocery*, planning time rebounds once the cache task count exceeds 30, showing that an overly large cache imposes additional overhead.

Low-workspace-overlap tasks include *Make Sandwich*, *Sort Cubs*, *Sweep Floor*, and *Arrange Cabinet*. As shown in the last four columns of Figure 4, success rates generally increase with larger

¹The appendix is available in the full version at <https://arxiv.org/abs/2601.20577>.

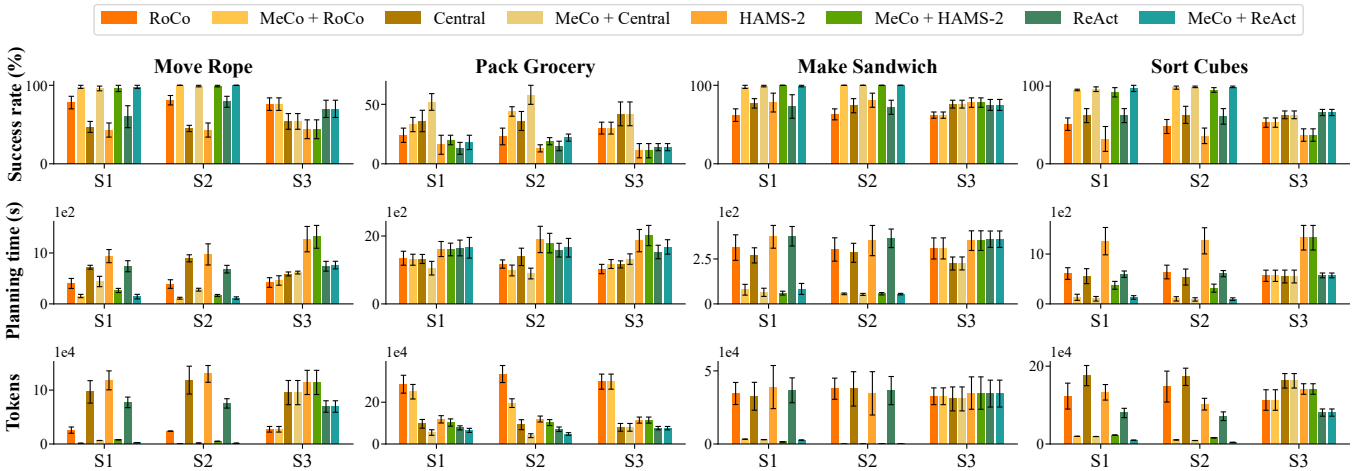


Figure 5: The performance of MeCo. We evaluate on MeCoBench across three scenarios: random (S1), totally similar (S2), and totally different (S3), using four baselines for comparison. For each task, we report success rate, planning time, and token consumption averaged over 30 random seeds. The results on *Sweep Floor* and *Arrange Cabinet* are provided in the Appendix C.1.

cache task count but converge at different rates across tasks. For example, *Make Sandwich* involves nearly 1,000 diverse scenarios, and its success rate approaches 100% only when the cache task count reaches about 500. In contrast, *Sort Cubes* stabilizes around a cache task count of 40, while *Sweep Floor* and *Arrange Cabinet*, which are inherently more similar, achieve near-perfect success rates when the cache task count is as small as 1. Unlike high-workspace-overlap tasks, the planning time for these tasks continues to decrease as the cache task count grows and eventually stabilizes, since the search overhead is negligible and does not cause a rebound. Meanwhile, token consumption consistently decreases with larger cache task count across all low-workspace-overlap tasks.

The results confirm the critical impact of cache task count on performance. For high-workspace-overlap tasks, an excessively large cache may increase search overhead, requiring a balance between leveraging task similarity and controlling search cost. For low-workspace-overlap tasks, however, a larger cache almost unilaterally improves performance, as the search overhead is negligible. Consequently, the optimal cache size varies across different task types. Moreover, MeCo allows the maximum cache size to be adjusted according to specific needs. A larger cache can reduce token consumption and improve success rates, but it may also result in longer planning times. Considering success rate, planning time, and token consumption together, we set the maximum cache sizes to 40, 30, 700, 50, 9, and 9 for the six tasks in subsequent experiments.

5.3 Performance of MeCo

We evaluate MeCo on MeCoBench under three scenarios: **random (S1)**, **totally similar (S2)**, and **totally different (S3)**. We use each of four baselines (RoCo [25], Central Plan, HMAS-2 [11], and ReAct [42]) as the LLM-empowered module within MeCo, and compare it with these strong baselines. As shown in Figure 5, MeCo significantly reduces token consumption while improving success rates and reducing planning time.

Success rate. The first row of Figure 5 presents the success rate performance of MeCo. In both the random (S1) and totally

similar (S2) scenarios, MeCo achieves a significant improvement over the four baselines across all tasks. Specifically, the average success rate increases by about 29% in the random scenario and 32% in the similar scenario, indicating that MeCo performs better as task similarity increases. The *Move Rope*, *Make Sandwich*, and *Sort Cubes* tasks nearly reach 100%, suggesting that when latent task-level similarity exist, MeCo effectively captures and transfers useful knowledge, achieving near-optimal performance. However, for the *Pack Grocery* task, all methods show relatively low success rates due to the high risk of collisions. Among them, ReAct and HMAS-2 reach only about 15%. Even so, MeCo still improves performance by roughly 8%, demonstrating its robustness in complex environments. Meanwhile, in the totally different (S3) scenario, MeCo performs on par with the baseline, indicating that it not only plans effectively under similar-task conditions but also maintains performance when tasks differ substantially.

Planning time. The second row of Figure 5 shows the reduction in planning time of MeCo. For the *Move Rope*, *Make Sandwich*, and *Sort Cubes* tasks, MeCo significantly reduces planning time in both the random (S1) and totally similar (S2) scenarios. Specifically, in the random scenario, the average reductions are 64% for *Move Rope*, 78% for *Make Sandwich*, and 76% for *Sort Cubes*. This shows that MeCo effectively reduces LLMs inference time by leveraging task-level similarities to decrease the number of LLMs calls. In contrast, the reduction is only 5% for *Pack Grocery*, mainly because this task involves a time-consuming search process, and the time saved by referencing similar tasks is less significant. In the similar scenario, the average reductions reach 77% for *Move Rope*, 12% for *Pack Grocery*, 83% for *Make Sandwich*, and 80% for *Sort Cubes*, indicating that higher task similarity leads to greater time savings. However, in the totally different (S3) scenario, the planning time increases by roughly 6%, suggesting that while MeCo introduces slight overhead in low-similarity conditions, the increase remains within an acceptable range.

Token usage. In the random (S1) scenario, MeCo significantly reduces token consumption compared with the baselines, achieving

Table 1: Ablation study on the similarity testing, S-Planner, and continuous planning module. The experiments are conducted under the MeCo which uses RoCo [25] as the LLM-empowered method. Evaluation results are averaged over 30 random seeds per task. Additional ablation studies on the selective caching mechanism are provided in the Appendix C.2.

		Move Rope	Pack Grocery	Make Sandwich	Sort Cubes	Sweep Floor	Arrange Cabinet
MeCo w/o Similarity testing	Success rate Time, Tokens	0.84 ± 0.05 ↓ 447.81s ↑, 17465 ↑	0.25 ± 0.06 ↓ 1573.88s ↑, 281165 ↑	0.90 ± 0.04 ↓ 128.30s ↑, 7679 ↑	0.78 ± 0.09 ↓ 291.46s ↑, 68592 ↑	0.99 ± 0.10 ≈ 75.86s ↑, 2164 ↑	1.00 ± 0 ≈ 43.58s ≈, 0 ≈
MeCo w/o S-Planner	Success rate Time, Tokens	0.78 ± 0.08 ↓ 444.35s ↑, 25737 ↑	0.24 ± 0.06 ↓ 1502.32s ↑, 286297 ↑	0.62 ± 0.08 ↓ 313.38s ↑, 34537 ↑	0.51 ± 0.11 ↓ 608.80s ↑, 123028 ↑	0.52 ± 0.09 ↓ 229.98s ↑, 23171 ↑	0.66 ± 0.12 ↓ 229.79s ↑, 32505 ↑
MeCo w/o Continuous Planning	Success rate Time, Tokens	0.97 ± 0.01 ↓ 214.29s ↑, 2871 ↑	0.30 ± 0.04 ↓ 1381.33s ↑, 261346 ↑	0.98 ± 0.02 ≈ 81.15s ↑, 3368 ↑	0.91 ± 0.03 ↓ 180.37s ↑, 51671 ↑	0.99 ± 0.10 ≈ 79.11s ↑, 1821 ↑	1.00 ± 0 ≈ 45.94s ≈, 0 ≈
MeCo	Success rate Time, Tokens	0.98 ± 0.02 154.63s, 1320	0.33 ± 0.06 1300.98s, 250264	0.98 ± 0.02 78.78s, 3304	0.95 ± 0.01 132.70s, 19684	0.99 ± 0.20 72.82s, 1443	1.00 ± 0 47.82s, 0

an average reduction of about 73% across all tasks. For the *Move Rope*, *Make Sandwich*, and *Sort Cubes* tasks, the reduction exceeds 90%, demonstrating the effectiveness of MeCo in lowering planning costs. In the totally similar (S2) scenario, the average token reduction reaches around 82%, as higher task similarity enables MeCo to make fewer LLM calls. However, *Pack Grocery* shows only an 18% reduction, mainly due to its inherently lower success rate. In the totally different (S3) scenario, token usage remains comparable to the baselines, indicating that MeCo does not introduce additional overhead when task similarity is low.

5.4 Ablation Studies

MeCo consists of several key components, including the similarity testing in the search stage, the similar motion planner (S-Planner), the continuous planning module, and the selective caching mechanism. To investigate the role of each design, we conduct an ablation study. Specifically, we select RoCo as the LLM-empowered method for this experiment. First, we remove the similarity testing and instead select the task with the smallest average positional change of objects as the similar task for compensation. Second, we remove the S-Planner and directly reuse the previous trajectory as a substitute. Finally, we remove the continuous planning module, when S-Planner fails, the planning process is restarted from scratch.

As shown in Table 1, after removing the similarity testing, the success rate drops significantly. Especially in the *Move Rope*, *Pack Grocery*, *Sort Cubes*, and *Make Sandwich* tasks, the performance decreases by about 12% on average. Meanwhile, both planning time and token consumption increase notably: *Move Rope* by 189%, *Pack Grocery* by 21%, *Make Sandwich* by 63%, and *Sort Cubes* by 120%. These results indicate that the absence of accurate similarity testing leads to ineffective task references and redundant reasoning, reducing overall efficiency. In contrast, for the *Sweep Floor* and *Arrange Cabinet* tasks, removing similarity testing causes no observable impact. This is because these tasks exhibit inherently high similarity, where even random task pairs share substantial resemblance, making complex filtering unnecessary. Overall, the results show that the similarity testing plays a crucial role in ensuring high-quality task search and preventing invalid matches.

Second, the performance degradation becomes more pronounced after removing the S-Planner. Since directly reusing previous trajectories results in nearly zero success rate, the overall success rate is

almost the same as that of directly using LLM-based planning, showing an approximate 30% drop compared with MeCo. Meanwhile, token consumption also aligns with that of direct LLM planning, exhibiting a significant increase. In particular, the planning time rises by 222.27%, even exceeding that of direct LLM planning, as additional time is required for matching similar tasks during the search stage. These results highlight the efficiency and superiority of the S-Planner in leveraging task-level similarity for planning.

Finally, after removing the continuous planning module, the success rate shows a slight decline in some tasks, while additional overhead is introduced, leading to a significant increase in both planning time and token consumption. This occurs because, without the continuous planning mechanism, LLMs must replan from scratch whenever the S-Planner fails locally, resulting in redundant reasoning and unnecessary computational cost. Moreover, the previous efforts of the S-Planner become wasted. Therefore, although this module has only a limited impact on the success rate, it plays a crucial role in improving overall reasoning efficiency.

6 CONCLUSION

In this paper, we propose MeCo, a similarity-aware multi-robot collaboration framework, that leverages task-level similarity for efficient planning. In the search stage, we categorize tasks into high-workspace-overlap and low-workspace-overlap types based on robots’ workspace overlap levels. Furthermore, we design a novel similarity testing method to identify previously solved tasks with high relevance, enabling the reuse of existing plans without re-invoking the LLMs. Within MeCo, the similar motion planner (S-Planner) efficiently generates plans by leveraging prior task plans. When S-Planner fails, MeCo activates the continuous planning module, which resumes planning from the failed step using LLMs. Once a task is successfully planned, MeCo selectively caches it in the cache for future reuse. Experiments show that our approach significantly reduces planning costs while improving success rates.

ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China (No. 2021YFB2900100), the National Science Fund for Distinguished Young Scholars (No. 62025205), and the National Natural Science Foundation of China (No. 62372383).

REFERENCES

- [1] How a universal model is helping one generation of Amazon robots train the next. <https://www.amazon.science/latest-news/how-a-universal-model-is-helping-one-generation-of-amazon-robots-train-the-next>. [Accessed 04-02-2025].
- [2] Pinch-grasping robot handles items with precision. <https://www.amazon.science/latest-news/pinch-grasping-robot-handles-items-with-precision>. [Accessed 04-02-2025].
- [3] Robot offline programming. <https://www.visualcomponents.com/products/robot-offline-programming/>. [Accessed 04-02-2025].
- [4] Shivendra Agrawal, Suresh Nayak, Ashutosh Naik, and Bradley Hayes. 2023. ShelfHelp: Empowering Humans to Perform Vision-Independent Manipulation Tasks with a Socially Assistive Robotic Cane. In *Proc. of AAMAS*.
- [5] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, et al. 2022. Do as I can, not as I say: Grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691* (2022).
- [6] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. In *Proc. of NeurIPS*.
- [7] Jiaqi Chen, Bingqian Lin, Xinmin Liu, Lin Ma, Xiaodan Liang, and Kwan-Yee K Wong. 2025. Affordances-oriented planning using foundation models for continuous vision-language navigation. In *Proc. of AAAI*.
- [8] 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 *Proc. of ICLR*.
- [9] Mengqi Chen, Bin Guo, Hao Wang, Haoyu Li, Qian Zhao, Jingqi Liu, Yasan Ding, Yan Pan, and Zhiwen Yu. 2025. The future of cognitive strategy-enhanced persuasive dialogue agents: new perspectives and trends. *Frontiers of Computer Science* (2025).
- [10] Weize Chen, Ziming You, Ran Li, Yitong Guan, Chen Qian, Chenyang Zhao, Cheng Yang, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. 2024. Internet of agents: Weaving a web of heterogeneous agents for collaborative intelligence. *arXiv preprint arXiv:2407.07061* (2024).
- [11] Yongchao Chen, Jacob Arkin, Yang Zhang, Nicholas Roy, and Chuchu Fan. 2024. Scalable multi-robot collaboration with large language models: Centralized or decentralized systems?. In *Proc. of IEEE ICRA*.
- [12] Yuanzhe Cui, Zhipeng Xu, Lou Zhong, Pengjie Xu, Yichao Shen, and Qirong Tang. 2024. A Task-Adaptive Deep Reinforcement Learning Framework for Dual-Arm Robot Manipulation. *IEEE TASAE* (2024).
- [13] Ken Goldberg. 2019. Robots and the return to collaborative intelligence. *Nature Machine Intelligence* (2019).
- [14] Jonathan Ho and Stefano Ermon. 2016. Generative adversarial imitation learning. In *Proc. of NeurIPS*.
- [15] Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, et al. 2023. Inner Monologue: Embodied Reasoning through Planning with Language Models. In *Proc. of CoRL*.
- [16] Jiabao Ji, Yongchao Chen, Yang Zhang, Ramana Rao Kompella, Chuchu Fan, Gaowen Liu, and Shiyu Chang. 2025. Collision-and Reachability-Aware Multi-Robot Control with Grounded LLM Planners. *arXiv preprint arXiv:2505.20573* (2025).
- [17] Hao Ju, Rongshun Juan, Randy Gomez, Keisuke Nakamura, and Guangliang Li. 2022. Transferring policy of deep reinforcement learning from simulation to reality for robotics. *Nature Machine Intelligence* (2022).
- [18] Matthew Lai, Keegan Go, Zhibin Li, Torsten Kröger, Stefan Schaal, Kelsey Allen, and Jonathan Scholz. 2025. RoboBallet: Planning for multirobot reaching with graph neural networks and reinforcement learning. *Science Robotics* (2025).
- [19] Audrey Lee, Turner S Baker, Joshua B Bederson, and Benjamin I Rapoport. 2024. Levels of autonomy in FDA-cleared surgical robots: a systematic review. *NPJ Digital Medicine* (2024).
- [20] Daniel Leidner, Wissam Bejjani, Alin Olimpiu Albu-Schäffer, and Michael Beetz. 2016. Robotic agents representing, reasoning, and executing wiping tasks for daily household chores. In *Proc. of AAMAS*.
- [21] Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. 2023. Code as policies: Language model programs for embodied control. In *Proc. of IEEE ICRA*.
- [22] Wenjie Lin, Jin Wei-Kocsis, Jiansong Zhang, Byung-Cheol Min, Dongming Gan, Paul Asunda, and Ragu Athinarayanan. 2025. Think, Reflect, Create: Metacognitive Learning for Zero-Shot Robotic Planning with LLMs. *arXiv preprint arXiv:2505.14899* (2025).
- [23] Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, et al. 2024. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437* (2024).
- [24] Jeffrey Mahler, Matthew Matl, Vishal Satish, Michael Danielczuk, Bill DeRose, Stephen McKinley, and Ken Goldberg. 2019. Learning ambidextrous robot grasping policies. *Science Robotics* (2019).
- [25] Zhao Mandi, Shreya Jain, and Shuran Song. 2024. Roco: Dialectic multi-robot collaboration with large language models. In *Proc. of IEEE ICRA*.
- [26] Qian Mao, Zijian Liao, Jinfeng Yuan, and Rong Zhu. 2024. Multimodal tactile sensing fused with vision for dexterous robotic housekeeping. *Nature Communications* (2024).
- [27] Ishaan Mehta, Keshav Bimbraw, Rajeevlochana G Chittawadigi, and Subir K Saha. 2016. A teach pendant to control virtual robots in Roboanalyzer. In *IEEE International Conference on Robotics and Automation for Humanitarian Applications*.
- [28] Ruaridh Mon-Williams, Gen Li, Ran Long, Wenqian Du, and Christopher G Lucas. 2025. Embodied large language models enable robots to complete complex tasks in unpredictable environments. *Nature Machine Intelligence* (2025).
- [29] Aritra Pal, Anandsingh Chauhan, and Mayank Baranwal. 2025. Together We Rise: Optimizing Real-Time Multi-Robot Task Allocation using Coordinated Heterogeneous Plays. In *Proc. of AAMAS*.
- [30] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* (2019).
- [31] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research* (2020).
- [32] Samuel Schmidgall, Ji Woong Kim, Alan Kuntz, Ahmed Ezzat Ghazi, and Axel Krieger. 2024. General-purpose foundation models for increased autonomy in robot-assisted surgery. *Nature Machine Intelligence* (2024).
- [33] Sangwon Seo and Vaibhav Unhelkar. 2025. Hierarchical Imitation Learning of Team Behavior from Heterogeneous Demonstrations. In *Proc. of AAMAS*.
- [34] Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. 2023. Progprompt: Generating situated robot task plans using large language models. In *Proc. of IEEE ICRA*.
- [35] Emanuel Todorov, Tom Erez, and Yuval Tassa. 2012. Mujoco: A physics engine for model-based control. In *Proc. of IEEE IROS*.
- [36] Zishen Wan, Yuhang Du, Mohamed Ibrahim, Jiayi Qian, Jason Jabbour, Yang Zhao, Tushar Krishna, Arijit Raychowdhury, and Vijay Janapa Reddi. 2025. ReCA: Integrated Acceleration for Real-Time and Efficient Cooperative Embodied Autonomous Agents. In *Proc. of ACM ASPLOS*.
- [37] Jun Wang, Guocheng He, and Yiannis Kantaros. 2024. Probabilistically Correct Language-Based Multi-Robot Planning Using Conformal Prediction. *IEEE Robotics and Automation Letters* (2024).
- [38] Eric Wete, Joel Greenyer, Daniel Kudenko, and Wolfgang Nejdl. 2024. Multi-Robot Motion and Task Planning in Automotive Production Using Controller-based Safe Reinforcement Learning. In *Proc. of AAMAS*.
- [39] Darryl L Willick, Derek L Eager, and Richard B Bunt. 1993. Disk cache replacement policies for network filesystems. In *Proc. of IEEE ICDCS*.
- [40] Lei Wu, Bin Guo, Qiuyun Zhang, Zhuo Sun, Jieyi Zhang, and Zhiwen Yu. 2023. Learning to Self-Reconfigure for Freeform Modular Robots via Altruism Proximal Policy Optimization. In *Proc. of IJCAI*.
- [41] Fan Xie, Alexander Chowdhury, M De Paolis Kaluza, Linfeng Zhao, Lawson Wong, and Rose Yu. 2020. Deep imitation learning for bimanual robotic manipulation. In *Proc. of NeurIPS*.
- [42] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models. In *Proc. of ICLR*.
- [43] Hongxin Zhang, Weihua Du, Jiaming Shan, Qinzhong Zhou, Yilun Du, Joshua B Tenenbaum, Tianmin Shu, and Chuang Gan. 2024. Building cooperative embodied agents modularly with large language models. In *Proc. of ICLR*.
- [44] 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).
- [45] Yang Zhang, Shixin Yang, Chenjia Bai, Fei Wu, Xiu Li, Zhen Wang, and Xuelong Li. 2025. Towards efficient llm grounding for embodied multi-agent collaboration. *Findings of ACL* (2025).
- [46] Zirui Zhao, Wee Sun Lee, and David Hsu. 2023. Large language models as commonsense knowledge for large-scale task planning. In *Proc. of NeurIPS*.