

# TREE-PLANNER: EFFICIENT CLOSE-LOOP TASK PLANNING WITH LARGE LANGUAGE MODELS

Mengkang Hu<sup>♣</sup> Yao Mu<sup>♣</sup> Xinmiao Yu<sup>♡</sup> Mingyu Ding<sup>\*♣</sup> Shiguang Wu<sup>◇</sup>  
Wenqi Shao<sup>♣</sup> Qiguang Chen<sup>♡</sup> Bin Wang<sup>◇</sup> Yu Qiao<sup>♣</sup> Ping Luo<sup>\*♣</sup>

## ABSTRACT

This paper studies close-loop task planning, which refers to the process of generating a sequence of skills (a plan) to accomplish a specific goal while adapting the plan based on real-time observations. Recently, prompting Large Language Models (LLMs) to generate actions iteratively has become a prevalent paradigm due to its superior performance and user-friendliness. However, this paradigm is plagued by two inefficiencies: high token consumption and redundant error correction, both of which hinder its scalability for large-scale testing and applications. To address these issues, we propose TREE-PLANNER, which reframes task planning with LLMs into three distinct phases: plan sampling, action tree construction, and grounded deciding. TREE-PLANNER starts by using an LLM to sample a set of potential plans before execution, followed by the aggregation of them to form an action tree. Finally, the LLM performs a top-down decision-making process on the tree, taking into account real-time environmental information. Experiments show that TREE-PLANNER achieves state-of-the-art performance while maintaining high efficiency. By decomposing LLM queries into a single plan-sampling call and multiple grounded-deciding calls, a considerable part of the prompt are less likely to be repeatedly consumed. As a result, token consumption is reduced by 92.2% compared to the previously best-performing model. Additionally, by enabling backtracking on the action tree as needed, the correction process becomes more flexible, leading to a 40.5% decrease in error corrections.

## 1 INTRODUCTION

Task planning is a significant topic in the field of robotics, where a system is tasked with crafting a sequence of mid-level actions (skills) that enable a robot to complete complex high-level tasks (Kaelbling & Lozano-Pérez, 2011). This involves a consideration of various factors, such as the capabilities of robots, the surrounding environment, and any constraints or uncertainties that might exist. An emerging trend within the field of task planning is using Large Language Models (LLMs) to generate actions directly (Huang et al., 2022a; Song et al., 2023), rather than searching within a pre-defined domain (Eysenbach et al., 2019; Xu et al., 2019).

As shown in Figure 1, the commonly adopted paradigm for LLM-based planning can be summarized as follows: (i) prompt an LLM to generate one action at a time; (ii) execute the generated action and then append the obtained observation to the LLM; and (iii) generate the next action. We categorize such approaches as ITERATIVE-PLANNER, which enables the model to generate subsequent actions in an auto-regressive manner. Based on ITERATIVE-PLANNER, when errors occur during action execution, existing research endeavors either re-generate actions at the current timestep (Raman et al., 2022; Guo et al., 2023) or re-generate the entire plan from the initial timestep (Shinn et al., 2023), referred to as LOCAL REPLAN and GLOBAL REPLAN, respectively.

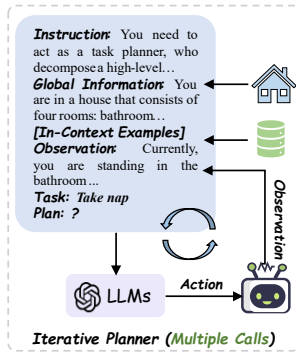


Figure 1: An overview of the traditional paradigm.

\*Corresponding authors: Mingyu Ding and Ping Luo ({dingmyu, pluo.lhi}@gmail.com). <sup>♣</sup>The University of Hong Kong. <sup>♡</sup>Harbin Institute of Technology. <sup>◇</sup>Noah’s Ark Laboratory. <sup>♣</sup>Shanghai AI Laboratory.

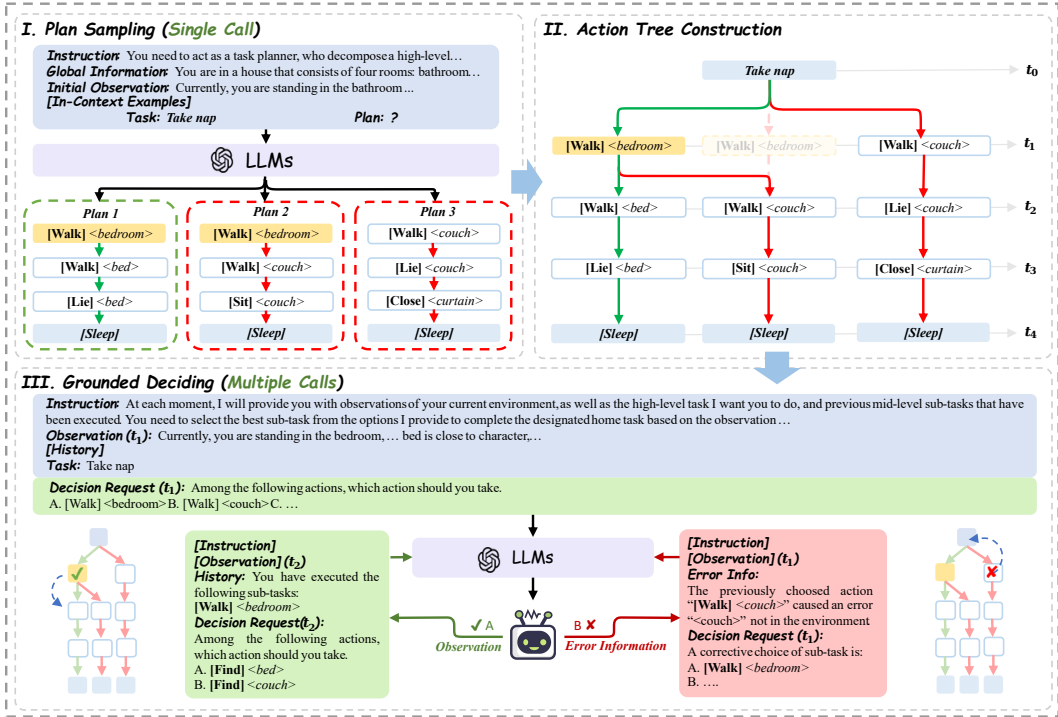


Figure 2: An overview of our TREE-PLANNER pipeline: (I) prompt an LLM to sample potential plans for “Take nap” before execution; (II) construct an action tree to aggregate sampled plans; (III) prompt the LLM again in closed loops to reason on the action tree. *Bottom-left*: “[WALK] <bedroom>” is successfully executed. Move on to the next level. *Bottom-right*: “[WALK] <couch>” fails because of the absence of “couch”. Mark the failed node as invalid, then track back and re-decide. The complete prompt and action tree can be found in Appendix F and Appendix G, respectively.

All methods above have the following two drawbacks: (i) **Token Inefficiency**: The expenses for a single LLM call increase proportionally with the number of tokens utilized, including both the *prompt tokens* and the *generated tokens*. However, in the scenario of task planning, the *prompt tokens* often consist of instructions, global information about the environment, in-context learning examples, and environmental observation (Vemprala et al., 2023) while the *generated tokens* predominantly represent a concise action. The discrepancy in the number of tokens between *prompt tokens* and *generated tokens* results in the issue of token inefficiency (Cheng et al., 2023). Moreover, due to the multi-step nature of a complex task (usually involving 5-20 steps), the *prompt tokens* incur repeated charges, leading to even higher costs. (ii) **Correction Inefficiency**: LOCAL REPLAN can be viewed as a trial-and-error approach implemented at the execution-failed time step, which makes it difficult for the model to detect errors that occurred several time steps earlier. While GLOBAL REPLAN can mitigate this problem by regenerating the entire plan, it may still come at the cost of increased time and token consumption. The token and correction inefficiencies inherent in ITERATIVE-PLANNER limit its applicability for large-scale inference or frequent use in everyday life.

To address the issues above while maintaining high performance, we propose TREE-PLANNER as illustrated in Figure 2. In general, TREE-PLANNER divides the queries to an LLM into two parts: a single plan-sampling call and multiple grounded-deciding calls to reduce the repetitive computational cost for several components in *prompt tokens*. These two stages are bridged using a tree-like structure, which leads to more effective logical correction. More specifically, TREE-PLANNER first prompts the LLM to sample potential task plans with its inherent commonsense (Stage I). Subsequently, an action tree is constructed to aggregate the sampled plans (Stage II). Lastly, TREE-PLANNER instructs the LLM again in closed loops to reason on the action tree with the environmental observations (Stage III). In terms of token efficiency, TREE-PLANNER only charges once for global information about the environment and in-context examples in plan sampling. However, for ITERATIVE-PLANNER, this information must be charged at each time step. In terms of correction efficiency, the correction process based on the action tree can be seen as an intermediate between

LOCAL REPLAN and GLOBAL REPLAN. TREE-PLANNER not only reduces the likelihood of redundant decision-making at a specific time step through backtracking but also significantly reduces the time and tokens required to generate the entire plan from scratch.

We demonstrate the effectiveness of TREE-PLANNER framework in VirtualHome (Puig et al., 2018), a simulated environment for complex household tasks. The experiments are conducted under two different settings: *with correction* and *without correction*. In *with correction* setting, the model is required to modify the plan when errors occur, while in *without correction* setting, the opposite is true. The main result shows that TREE-PLANNER achieves state-of-the-art results in both experimental settings, surpassing the best baseline models by 1.29% and 3.65% in terms of success rate, respectively. At the same time, TREE-PLANNER exhibits high efficiency. In terms of token efficiency, TREE-PLANNER reduces the token cost of ITERATIVE-PLANNER by 53.29%. Furthermore, when compared to LOCAL REPLAN and GLOBAL REPLAN under the *with correction* setting, TREE-PLANNER achieves even greater improvement with reductions of 74.36% and 92.24%, respectively. In terms of correction efficiency, TREE-PLANNER reduces the number of corrections by 37.99% and 40.52%, respectively. In further analysis, we formally verify the token efficiency of TREE-PLANNER and derive the critical value of the number of sampled plans required for the model to possess token efficiency. We also perform an ablation study on both plan sampling and grounded deciding, demonstrating the effectiveness of the individual components of TREE-PLANNER. Finally, we provide a manual error analysis of potential areas for improvement in the model.

## 2 PRELIMINARY

**Task and Motion Planning (TAMP)** (Kaelbling & Lozano-Pérez, 2011) is the process of generating a sequence of actions and robot motions to achieve a desired goal in a given environment. As is shown in Figure 2, a high-level task description such as “Take nap” is decomposed into several mid-level actions. We assume the existence of a low-level controller that can execute these mid-level actions, which typically requires training using reinforcement learning (RL) methods or fine-tuning with expert data. Task planning can be categorized into closed-loop task planning and open-loop task planning. Open-loop task planning aims to decompose a high-level task description into a mid-level plan without any feedback from the environment. Closed-loop task planning, on the other hand, involves continuously adjusting planning strategies through perception and feedback mechanisms to adapt to environmental changes and uncertainties during execution. This paper focuses on closed-loop task planning, which is more suitable for task execution in dynamic and complex environments.

**Problem Setup** We formulate the closed-loop task planning problem as a partially observable Markov decision processes (POMDPs) denoted by  $\langle S, O, A, \mathcal{T} \rangle$ , which is similar to Li et al. (2022a).  $S, O, A$  are sets of states, observations and actions respectively and  $\mathcal{T}(s_{t+1}|s_t, a_t)$  is a transition model. In a POMDP setting, the observation  $o_t$  represents a subset of the underlying state  $s_t$ . Let  $g$  be the task, the optimal policy  $\pi(a_t|g, h_t, o_t)$  must take into account not only the current observation  $o_t$ , but also the entire history of actions  $h_t = \{a_1, \dots, a_{t-1}\}$ .

## 3 MODEL

### 3.1 PLAN SAMPLING

Abstract specifications often restrict task planning. Take the “Take nap” task as an example, the robot needs to understand that napping can be done on a bed, and the bed is typically located in a bedroom. Many works hold the belief that LLMs trained on large-scale data encode commonsense knowledge about the real-world (Davison et al., 2019; Li et al., 2022b; Bian et al., 2023). Recently, several studies have investigated the integration of LLMs into task planning, which aims to address language ambiguities and provide robots with background knowledge (Huang et al., 2022a; Li et al., 2022a; Ahn et al., 2022). In contrast to these approaches, which typically use LLMs directly as policies, TREE-PLANNER prompts an LLM to generate prospective task plans before executing them in a specific environment. We consider this as a way to extract commonsense knowledge from LLM through sampling, which serves as prior knowledge for task planning. Let  $\rho_{ps}$  be the prompt for plan sampling,  $g$  be the task name, the process of plan sampling can be formalized as:  $\text{LLM}(\rho_{ps}, g) = \mathbf{c} = \{c_1, c_2, \dots, c_N\}$ , where  $N$  is a hyper-parameter which determines the number of sampled

plans. Each plan candidate  $c_i$  is a sequence of actions, i.e.,  $c_i = \{a_{it} | t = 1, \dots, m(i)\}$ .  $m(i)$  is the number of actions in plan  $i$  and  $a_{it}$  is the action of plan  $i$  at time step  $t$ . The prompt consists of four parts: *instruction*, *global information*, *initial observation*, and *in-context examples*. The *instruction* provides the LLM with a clear and concise explanation of the process of task planning. The *global information* provides the LLM with background knowledge about the environment and available action space. The *initial observation* provides the LLM with an initial snapshot at the starting point of the task. The *in-context examples* are additional task plans that serve to indicate the format of the output plan and have also been proven to be helpful in enhancing performance (Brown et al., 2020). In Section 5.2, we provide a quantitative analysis of the upper-bound on plan sampling.

### 3.2 ACTION TREE CONSTRUCTION

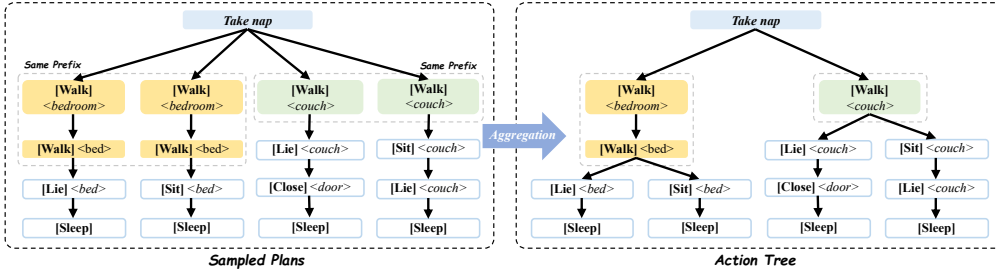


Figure 3: The process of constructing the action tree. *Left*: each path represents a sampled plan. *Right*: plans with the same prefix are aggregated together. Note that although certain paths have the same action ([Sleep]), they are not aggregated together due to inconsistent prefixes.

To select an optimal plan from potential plans, an obvious approach would be to execute and test each plan in the environment. However, this approach has two drawbacks: **(i)** It is time-consuming to execute multiple plans in the environment; **(ii)** Different plans may have overlapping parts, so repeating the execution of these overlapping parts in the environment is redundant. For example, in *plan 1* and *plan 2* shown in Figure 2, the first step in both plans is: “[Walk] <bedroom>”. Based on the previous analysis, we designed a structured representation that aggregates the sampled potential plans called Action Tree. As is shown in Figure 3, when two plans share a common prefix but differ in their actions at a specific time step, their shared prefix is aggregated into a single branch, while their differing actions form divergent paths. This process repeats iteratively until all sampled plans are organized into a complete tree structure. The motivation behind it is to convert the filtering of the plan level into a search at the action level, thereby reducing the execution time in the environment. an action tree with root node  $r$  can be formalized as  $T(c) = (V, E)$ , where  $V$  and  $E$  represent the sets of nodes and edges respectively. Each node  $v$  is associated with an action  $a_v$  and a time step  $t_v$ , i.e.,  $v = (a_v, t_v)$ . Each edge  $e$  represents a pair of adjacent actions in plan  $c_i$ , i.e.,  $E = \{e(v_1, v_2) | v_1, v_2 \in V, v_1 = (a_{it}, t), v_2 = (a_{i(t+1)}, t + 1)\}$ . The root node  $r$  is not associated with any specific action, and its child nodes are the set of nodes obtained by aggregating the first action of each plan. The construction process of the action tree is presented in Algorithm 1.

### 3.3 GROUNDED DECIDING

During grounded deciding, an LLM functions as the policy  $\pi(a_t | g, h_t, o_t)$ . However, instead of sampling from the entire corpus of LLMs as the ITERATIVE-PLANNER, we limit the choices to a few child nodes of the current node at time  $t$  on the action tree. This process simulates the decision-making process of humans, who first propose several action options and then combine their current real-world observations to make decisions. Specifically, we provide an LLM with *instruction*, *observation*, and *history* (the previously executed actions) as prompts, and then the LLM chooses one from the child nodes of the current node. Furthermore, we also designed a corresponding error correction method. When a chosen action fails to execute in the environment, TREE-PLANNER **(i)** marks the nodes on the subtree rooted at the failed node as invalid nodes; **(ii)** traces back on the action tree to find the previous valid fork node with available valid child nodes. If all the child nodes of a particular node are invalid, then the fork node should also be marked as invalid. **(iii)** executes the inverse process of previously executed actions (e.g., the inverse of [SwitchOn] is [SwitchOff]) to recover the state of the agent; **(iv)** re-decides at the fork node. Error correction with grounded

**Algorithm 1:** Action Tree Construction

---

```

Input :  $c, r$ 
Output:  $r$ 
1 Function ConstructActionTree( $c, r$ ):
2   forall  $c_i \in c$  do
3      $n \leftarrow r$ ;
4     for  $t = 1$  to  $m(i)$  do
5        $cn \leftarrow \text{GetChildNode}(n, a_{it})$ ;
6       if  $cn$  is None then
7          $cn \leftarrow \text{CreateChildNode}(a_{it})$ ;
8          $\text{AddChildNode}(n, cn)$ ;
9       end
10       $n \leftarrow cn$ ;
11    end
12  end

```

---

deciding is more effective than the commonly adopted methods presented in Section 1. This is because the action tree serves as an important prior to completing the current task. Therefore, when an error occurs at a node on the tree, it is possible to selectively backtrack on the action tree, thus alleviating repetitive decisions at a particular time step as in LOCAL REPLAN. Performing error correction on the action tree also relieves the need to return to the initial time step as in GLOBAL REPLAN, thereby reducing time and token consumption. The process described above is displayed in Figure 4. Quantitative analysis of the effectiveness of error correction is presented in Section 5.3.

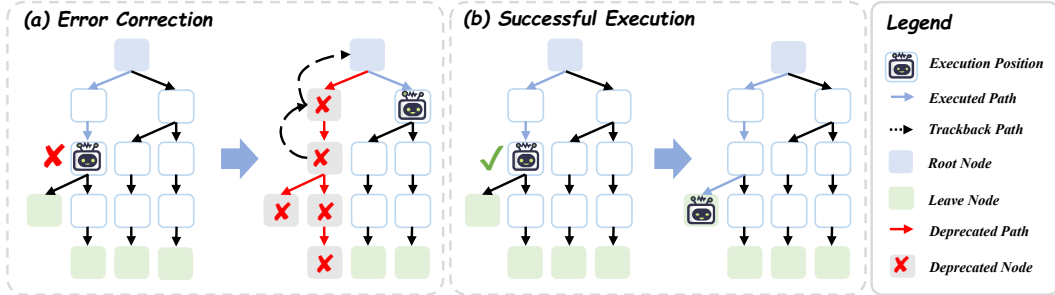


Figure 4: An overview of the process of *grounded deciding*. *Left*: When an error occurs, TREE-PLANNER tracks back and marks the nodes along the way as invalid. Afterward, TREE-PLANNER makes a new decision at the previous fork node. *Right*: After the action is successfully executed, TREE-PLANNER makes a decision at the current node, and then the agent moves on to the next level.

## 4 EXPERIMENTAL RESULTS

### 4.1 EXPERIMENTAL SETUP

**Environment.** We conduct the experiments in the VirtualHome (VH) Environment (Puig et al., 2018), a simulation platform for household tasks. Each scene in every VH environment contains hundreds of objects. These objects may possess individual properties, and there may also be relationships between different objects. There are 28 different action types in VH, which are listed in Appendix A.1. The task-relevant goal conditions refer to a set of specific states of objects or predicates between objects. For example, a goal condition for *Turn on TV* would be  $On(TV)$ , while a goal condition for *Sit* would be  $On(character, chair)$ .

**Dataset.** We constructed a dataset consisting of 4 VH scenes and 35 unique VH tasks. Each task includes a task name, goal conditions, and a gold plan. We started by annotating goal conditions for each task from *ActivityPrograms* knowledge base by Puig et al. (2018) via executing the programs. And then, we applied simple heuristics to filter the low-quality annotations in the dataset: (i) the length of the plan is less than 3; (ii) the execution of the program fails. To highlight the necessity of grounding LLMs in the real environment which has variation in the objects and preconditions, we

replicated the annotation above process across 4 distinct scenes provided in VirtualHome, ultimately yielding 71 annotated tasks. We denote the 4 distinct scenes as ENV- $\{1, 2, 3, 4\}$ . Then, we hired two CS-majored graduate students to conduct manual quality control to ensure that the task descriptions were in line with their corresponding goal conditions and programs. We eliminate cases that do not meet the alignment criteria or were originally annotated with errors, resulting in a high-quality dataset comprising 35 tasks. To double-check the quality of the dataset, we also study the agreement between annotators. The results indicated “almost perfect agreement” with Fleiss Kappa (Landis & Koch, 1977) scores of 0.88.

**Evaluation Metrics.** We use four metrics to evaluate the performance of different methods: executability (EXEC.), success rate (SR), goal conditions recall (GCR), and the financial expenditure for evaluation (\$COST). EXEC. refers to whether the plan can be executed in the given environment, regardless of its relevance to the task. GCR is calculated by taking the difference between the ground truth goal conditions and the goal conditions that were achieved with the generated plan and then dividing this difference by the total number of goal conditions. SR measures whether all goal conditions are fulfilled, i.e.,  $SR = 1$  only when  $GCR = 1$ . \$COST is used to evaluate the token efficiency of different methods, which is calculated based on the pricing provided by OpenAI.<sup>1</sup> For evaluation with error correction, we use NO.EC to represent the number of error corrections of each method. NO.EC does not directly measure performance but rather evaluates how effectively different models can correct errors.

**Baselines.** For experiments without error correction, we compare our method to two strong published LLM-based task planning methods with OpenAI APIs, including: (i) ZERO-SHOT PLANNER (Huang et al., 2022a); (ii) PROG PROMPT (Singh et al., 2022). Furthermore, we also implement the ITERATIVE-PLANNER method discussed in Section 1 as a baseline model. For experiments with error correction, we enhance the ITERATIVE-PLANNER method with the two re-planning methods: LOCAL REPLAN and GLOBAL REPLAN, and consider them as the baseline models. More implementation details and an introduction to each baseline model can be found in Appendix B.2.

**Implementation Details.** We use the OpenAI GPT-3.5 (text-davinci-003) API<sup>2</sup> model as a LLM backbone in our experiments for all evaluated methods. The cost of this model is 0.02\$ per 1000 tokens. The prompt for TREE-PLANNER and ITERATIVE-PLANNER was designed with the principles proposed in Vemprala et al. (2023), and examples can be found in Appendix F. We take 4 representative tasks from the dataset as in-context learning exemplars and the rest as the validation set. The examples are fixed to be: “Watch TV”, “Turn on light”, “Go to sleep”, and “Brush teeth”. To sample diverse plans, we applied a temperature of 0.8 and a top-p value of 0.95. We heuristically set the number of samplings  $N \in \{25, 50\}$ . During grounded deciding, we set the temperature to 0.7, top-p to 1.0, and sampling parameter n to 20. Additionally, we utilize a majority vote to obtain the final option in order to alleviate format errors in the output of LLMs. The maximum number of error corrections is set to 10 for all evaluated approaches.

## 4.2 MAIN RESULTS

Based on the results presented in Table 1, several advantages of TREE-PLANNER can be derived: (i) TREE-PLANNER outperforms listed baseline systems, surpassing the previous state-of-the-art by absolute 11.2% and 7.04% on Executability, 6.71% and 7.29% on GCR and 1.29% and 3.65% on SR under both experimental settings respectively. This experimental observation demonstrates that reframing the LLM-based planning pipeline does not compromise its performance. (ii) TREE-PLANNER has a significant advantage in token efficiency. In *without correction* setting, TREE-PLANNER reduces the cost of ITERATIVE-PLANNER by 53.29%. In *with correction* setting, the token cost is further reduced by 74.36% and 92.24%, respectively, compared to LOCAL REPLAN and GLOBAL REPLAN. (iii) TREE-PLANNER also demonstrates high correction efficiency, resulting in a reduction of the number of action-retry times for LOCAL REPLAN and GLOBAL REPLAN by 37.99% and 40.52%, respectively. A reduced amount of corrections also contributes to a decrease in token consumption.

Note that, while not having a token efficiency advantage compared to ZERO-SHOT PLANNER and PROG PROMPT, TREE-PLANNER significantly outperforms these methods in terms of performance

<sup>1</sup><https://openai.com/pricing>

<sup>2</sup><https://openai.com/>

Table 1: Performance of different methods on Virtual Home. *w/o correction* means that during the plan execution, there is no allowance for retrying failed actions. While *with correction* implies the opposite. The reported evaluation metrics are the average of 3 independent runs across the 4 scenes.

	EXEC. $\uparrow$	SR $\uparrow$	GCR $\uparrow$	\$Cost $\downarrow$	No.EC $\downarrow$
<b><i>w/o correction</i></b>					
ZERO-SHOT PLANNER	16.49 $\pm$ 3.08	1.07 $\pm$ 0.76	1.52 $\pm$ 0.75	1.36 $\pm$ 0.09	N/A
PROGPROMPT	35.04 $\pm$ 3.98	12.54 $\pm$ 2.20	19.99 $\pm$ 2.83	<b>1.25</b> $\pm$ 0.55	N/A
ITERATIVE-PLANNER	44.54 $\pm$ 6.09	27.04 $\pm$ 4.65	33.25 $\pm$ 5.32	5.12 $\pm$ 0.14	N/A
TREE-PLANNER <sub>N=25</sub>	<b>55.74</b> $\pm$ 0.92	<b>28.33</b> $\pm$ 1.18	<b>39.96</b> $\pm$ 0.16	2.39 $\pm$ 0.44	N/A
TREE-PLANNER <sub>N=50</sub>	49.01 $\pm$ 5.67	28.14 $\pm$ 2.45	35.84 $\pm$ 4.20	3.48 $\pm$ 0.04	N/A
<b><i>with correction</i></b>					
LOCAL REPLAN	79.66 $\pm$ 2.33	37.46 $\pm$ 1.71	51.9 $\pm$ 0.15	12.88 $\pm$ 0.17	3.29 $\pm$ 0.46
GLOBAL REPLAN	82.09 $\pm$ 1.32	37.93 $\pm$ 1.22	52.46 $\pm$ 0.86	42.55 $\pm$ 0.09	3.43 $\pm$ 0.15
TREE-PLANNER <sub>N=25</sub>	<b>89.13</b> $\pm$ 0.17	35.30 $\pm$ 1.78	56.65 $\pm$ 1.09	<b>3.30</b> $\pm$ 0.01	<b>1.85</b> $\pm$ 0.05
TREE-PLANNER <sub>N=50</sub>	88.26 $\pm$ 2.47	<b>41.58</b> $\pm$ 3.20	<b>59.55</b> $\pm$ 3.20	4.54 $\pm$ 0.16	2.04 $\pm$ 0.26

by 27.26% and 15.79% on SR respectively. It is also worth noting that increasing the hyper-parameter  $N$  does not result in consistently improved performance. This experimental phenomenon will be further discussed in Section 5.2.

## 5 ANALYSIS

### 5.1 TOKEN EFFICIENCY

In Section 4.2, the quantitative analysis has demonstrated that TREE-PLANNER consumes fewer tokens compared to ITERATIVE-PLANNER. In this section, we will further provide a specific formulation to demonstrate this point. The number of tokens required for an LLM API call typically includes two parts: *prompt tokens* and *generated tokens*. Let  $\rho$  and  $\varphi$  represent the prompt tokens and generated tokens. Let  $ps$ ,  $gd$ ,  $ip$  stand for plan sampling, grounded deciding, and ITERATIVE-PLANNER, respectively. Normally, we have  $\rho_{ip} \approx \rho_{ps} + \rho_{gd}$ . That is because, as shown in Figure 2 and Figure 1, the prompt for plan sampling typically includes global information and in-context examples, while the prompt for grounded deciding includes observation and history. These types of information usually need to be included in every step of ITERATIVE-PLANNER. Assuming that the number of tokens for each action type  $|a|$  is the same and the total number of steps  $M$  is the same for each generated plan. The hyper-parameter number of sampling is  $N$  for plan sampling and grounded decoding and is 1 for ITERATIVE-PLANNER. Based on the given information, we have  $\varphi_{ps} = MN|a|$ ,  $\varphi_{gd} = N$  and  $\varphi_{ip} = |a|$ . The consumed tokens  $\mu_{ours}$  and  $\mu_{ip}$  can be calculated as follows:  $\mu_{ours} = \rho_{ps} + \varphi_{ps} + M \cdot (\rho_{gd} + \varphi_{gd})$  and  $\mu_{ip} = M \cdot (\rho_{ip} + \varphi_{ip})$ . Based on the above formula, we can determine the boundary conditions for  $N$  that satisfy the inequality  $\mu_{ours} < \mu_{ip}$  as follows:  $N < \frac{1-1/M}{1+1/|a|} \cdot \frac{\rho_{ps}}{|a|} + \frac{|a|}{|a|+1}$ . And we have  $\rho_{ps} \gg |a|$ , since the prompt of plan sampling may contain thousands of tokens and an action only contains a few tokens. We use the average number of tokens for all action types to estimate  $|a|$  and the average length of all gold plans to estimate  $M$ . As a result, we obtain the critical value of  $N$  in our experiment as follows:  $N < 197.72$ . Detailed derivation can be found in Appendix D. In conclusion, our model exhibits a remarkably high token efficiency, especially in scenarios where  $N$  is not particularly high.

### 5.2 PLAN SAMPLING

Since grounded deciding fundamentally involves selecting from the sampled plans, the upper limit of our TREE-PLANNER is determined by plan sampling. We propose two additional metrics to study the upper limit of plan sampling: **(i)** the maximum GCR for all generated plans, i.e.,  $GCR_{max}(e) = \max_{i=1}^N(GCR(c_i))$ ; **(ii)** the average GCR for all generated plans, i.e.,  $GCR_{avg}(e) = \frac{1}{N} \sum_{i=1}^N(GCR(c_i))$ .  $GCR_{max}$  represents the upper limit of the performance of TREE-PLANNER. In other words, the model can only succeed if there is a “correct” plan among the sampled plans.  $GCR_{avg}$  reflects the proportion of “correct” plans to sampled plans. When  $GCR_{avg}$  is low, it undoubtedly poses greater

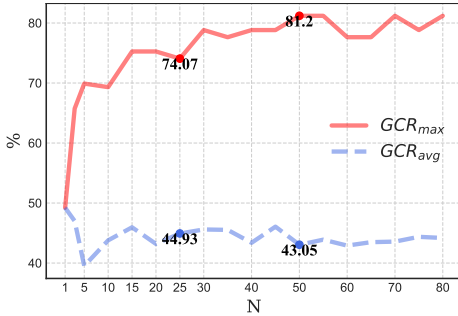


Figure 5: Maximum and average GCR for all sampled plans. The x-axis of the annotated coordinate points represents the chosen  $N$  for the main experiments.

challenges for grounded deciding. Some conclusions can be drawn from Figure 5: **(i)** The maximum value of  $GCR_{max}$  being 81.2% indicates that plan sampling is effective. **(ii)** As  $N$  increases, there is a noticeable increase in  $GCR_{max}$ , but it eventually reaches a threshold. Therefore, a large value of  $N$  will lead to increased token consumption without necessarily improving the performance limit. When applying TREE-PLANNER, it is essential to choose an appropriate value of  $N$  that balances token assumption and model performance. **(iii)**  $GCR_{avg}$  does not consistently increase with an increased  $N$ . This implies that as  $N$  becomes larger, the proportion of “correct” plans to sampled plans may not necessarily increase.

### 5.3 GROUNDED DECIDING

To investigate the effectiveness of grounded deciding, we conducted ablation experiments. We incorporated the gold plan for each task into the construction of the action tree. As is shown in Table 2, after incorporating the gold plan, there was a significant improvement in performance. Additionally, there was also a decrease in the number of error corrections. For TREE-PLANNER $_{N=25}$ , the number decreased from 1.85 to 1.21, and for TREE-PLANNER $_{N=50}$ , it decreased from 2.04 to 1.39. The quantitative experimental results presented above demonstrate the effectiveness of grounded deciding. Another noteworthy experimental phenomenon is that the improvement in performance for TREE-PLANNER $_{N=25}$  was greater than that for TREE-PLANNER $_{N=50}$ . This further validates the conclusion we drew in Section 5.2: when the number of plans increases, but the proportion of correct plans decreases, the performance may be negatively impacted.

### 5.4 ERROR ANALYSIS

We categorize error types into two distinct classifications: **(i)** Missing Correct Plan; **(ii)** Grounded Deciding Error. As is listed in Table 3, the majority of errors are attributed to the missing correct plans (54.5%). Therefore, despite the ability of plan sampling to achieve relatively high  $GCR_{max}$  as is discussed in Section 5.2, it still serves as a bottleneck for our model to some extent. Furthermore, a considerable portion of the errors occurred due to mistakes made by LLM during grounded deciding (45.5%). We also provide a qualitative analysis of each error type in Appendix E.

Table 3: Distribution of error types of the TREE-PLANNER $_{N=25}$  *w/o correction* model.

Error Type	Explanation	Proportion(%)
Missing Correct Plans	Plan sampling did not yield correct plans	<b>54.5%</b>
Environment Misunderstanding	Misunderstandings on actions or objects	18.2%
Incomplete Plan	The absence of essential steps	18.2%
Illogical Error	The generated plan is logically incorrect	13.6%
Semantically Correct	Execution failed but semantically correct	9.1%
Grounded Deciding Error	Execution error during grounded deciding	<b>45.5%</b>
Incorrect Deciding	Incorrect decisions at specific nodes	31.8%
Semantically Correct	Execution failed but semantically correct	13.7%

	EXEC.	SR	GCR
<i>w/o correction</i>			
TREE-PLANNER $_{N=25}$	55.74	28.33	38.96
† with oracle †	7.16	9.84	8.5
TREE-PLANNER $_{N=50}$	49.01	28.14	35.84
† with oracle †	3.41	6.54	4.78
<i>with correction</i>			
TREE-PLANNER $_{N=25}$	89.13	35.3	56.65
† with oracle †	8.45	26.8	19.76
TREE-PLANNER $_{N=50}$	88.26	41.58	59.55
† with oracle †	6.9	10.57	7.47

Table 2: Ablation study on grounded deciding. † represents the performance improvement after adding a gold plan to action tree construction.



## 6 RELATED WORKS

**Task Planning with Large Language Models.** We categorize the mainstream methods in the task planning domain into two groups: search-based methods (Jiang et al., 2018; Garrett et al., 2018) and generate-based methods (Song et al., 2023; Wu et al., 2023a; Ding et al., 2023; Mu et al., 2023). LLMs trained on a large-scale corpus contain a vast amount of commonsense knowledge for task planning (Pallagani et al., 2023; Sun et al., 2023b;a). Thanks to this advancement, generate-based methods have gradually become a hot topic of research in recent years. When considering the utilization of LLM, some works directly generate the entire plan without executing in the environment (Singh et al., 2022; Liang et al., 2023; Wu et al., 2023b; Zeng et al., 2023; Lin et al., 2023b; Yang et al., 2023). While these models possess token efficiency, they are unable to modify the plan when encountering errors dynamically. Another line of works has adopted the paradigm presented in Section 1 to generate actions iteratively (Vemprala et al., 2023; Yao et al., 2022; Huang et al., 2022a;b; Shinn et al., 2023), which is more flexible for error correction, human interaction and the grounding of environment. Works like Carta et al. (2023); Huang et al. (2023); Ahn et al. (2022) involve the use of implicit representations of LLM. In contrast to these works, our study concentrates on Black-box LLMs, which are utilized in a manner more frequently by researchers and industry, as they provide only input and output without any additional information.

**Tree-based Modeling for the Output of Large Language Models.** Yao et al. (2023); Long (2023) both propose an alternative for chain-of-thought, called “tree-of-thought”, for problem-solving. These studies do not involve the interaction between inner steps in the tree and the environment but rather focus on reasoning tasks. Considering the robotic area, Cao & Lee (2023) leverages LLMs for automatic behavior-tree-based task generation. Zhao et al. (2023); Hao et al. (2023) propose using an LLM as a world model to assist planning algorithms such as Monte Carlo Tree Search (MCTS). However, TREE-PLANNER samples diverse paths once and aggregates the paths into an action tree rather than requiring multiple calls to LLM like the aforementioned studies. This approach offers advantages in terms of both run-time efficiency and token efficiency.

**Generate then Select.** From another perspective, grounded deciding selects a prediction from the sampled potential plans. Hence, TREE-PLANNER follows the paradigm of *generate then select*, which is commonly adopted to optimize the output of LLMs. Some models (Glass et al., 2022; Suzgun et al., 2022; Wang et al., 2023b; Gu et al., 2023) use external controllers to re-rank the generations. In Wang et al. (2023a), the best answer is selected from multiple generations of an LLM through a majority vote. Logeswaran et al. (2022) proposes to incorporate the state information from the environment to re-rank the generated plans. Unlike these works, instead of selecting at the level of entire generation, we use action trees to perform more fine-grained selection (action-level).

**Efficient Inference with Large Language Models.** Most previous works suggest modifying the architecture of transformer or decoding strategy to achieve efficient inference (Wang et al., 2020; Katharopoulos et al., 2020; Leviathan et al., 2023; Chen et al., 2023). Cheng et al. (2023) propose a batch prompting method to reduce the frequency of invoking LLMs. Lin et al. (2023a) achieve efficient inference with LLMs by incorporating a small LM fine-tuned on oracle trajectories. TREE-PLANNER differs from previous studies by simply reframing the process of LLM planning to alleviate repeated token consumption without the need for additional training.

## 7 CONCLUSION

In this paper, we have introduced TREE-PLANNER, a novel framework for task planning with LLMs. The motivation behind TREE-PLANNER is to address the inefficiencies of the commonly adopted paradigm while still achieving high performance. Through extensive experiments in the Virtual-Home environment, we have demonstrated that TREE-PLANNER outperforms other strong baselines and achieves state-of-the-art performance. We have also shown that our framework is highly efficient in terms of token consumption and error correction. To gain a deeper understanding of our framework, we have conducted several studies analyzing its performance gains and identifying potential bottlenecks. Furthermore, we have performed a qualitative error analysis to identify areas where the model may fail. Overall, we believe that TREE-PLANNER represents a new paradigm for LLM-based task planning that strikes a balance between efficiency and performance. We hope that our work will inspire further research and the development of more efficient task-planning methods.

## 8 ETHICS STATEMENTS

We build the dataset based on the *ActivityPrograms* knowledge base by Puig et al. (2018), which is under the MIT license. Our approach has no ethical or social issues on its own, except those inherited from large language models.

## 9 ACKNOWLEDGMENTS

This paper is partially supported by the National Key R&D Program of China No.2022ZD0161000 and the General Research Fund of Hong Kong No.17200622.

## REFERENCES

- Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Daniel Ho, Jasmine Hsu, Julian Ibarz, Brian Ichter, Alex Irpan, Eric Jang, Rosario Jauregui Ruano, Kyle Jeffrey, Sally Jesmonth, Nikhil J Joshi, Ryan Julian, Dmitry Kalashnikov, Yuheng Kuang, Kuang-Huei Lee, Sergey Levine, Yao Lu, Linda Luu, Carolina Parada, Peter Pastor, Jornell Quiambao, Kanishka Rao, Jarek Rettinghouse, Diego Reyes, Pierre Sermanet, Nicolas Sievers, Clayton Tan, Alexander Toshev, Vincent Vanhoucke, Fei Xia, Ted Xiao, Peng Xu, Sichun Xu, Mengyuan Yan, and Andy Zeng. Do as i can, not as i say: Grounding language in robotic affordances, 2022. 3, 9
- Ning Bian, Xianpei Han, Le Sun, Hongyu Lin, Yaojie Lu, and Ben He. Chatgpt is a knowledgeable but inexperienced solver: An investigation of commonsense problem in large language models, 2023. 3
- Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020. 4
- Yue Cao and C. S. George Lee. Robot behavior-tree-based task generation with large language models, 2023. 9
- Thomas Carta, Clément Romac, Thomas Wolf, Sylvain Lamprier, Olivier Sigaud, and Pierre-Yves Oudeyer. Grounding large language models in interactive environments with online reinforcement learning, 2023. 9
- Charlie Chen, Sebastian Borgeaud, Geoffrey Irving, Jean-Baptiste Lespiau, Laurent Sifre, and John Jumper. Accelerating large language model decoding with speculative sampling, 2023. 9
- Zhoujun Cheng, Jungo Kasai, and Tao Yu. Batch prompting: Efficient inference with large language model apis, 2023. 2, 9
- Joe Davison, Joshua Feldman, and Alexander Rush. Commonsense knowledge mining from pre-trained models. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 1173–1178, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1109. URL <https://aclanthology.org/D19-1109>. 3
- Yan Ding, Xiaohan Zhang, Chris Paxton, and Shiqi Zhang. Task and motion planning with large language models for object rearrangement, 2023. 9
- Benjamin Eysenbach, Ruslan Salakhutdinov, and Sergey Levine. Search on the replay buffer: Bridging planning and reinforcement learning. *arXiv: Artificial Intelligence, arXiv: Artificial Intelligence*, Jun 2019. 1

- Caelan Reed Garrett, Tomás Lozano-Pérez, and Leslie Pack Kaelbling. Pddlstream: Integrating symbolic planners and blackbox samplers via optimistic adaptive planning. *arXiv: Artificial Intelligence, arXiv: Artificial Intelligence*, Feb 2018. 9
- Michael Glass, Gaetano Rossiello, Md Faisal Mahbub Chowdhury, Ankita Rajaram Naik, Pengshan Cai, and Alfio Gliozzo. Re2g: Retrieve, rerank, generate, 2022. 9
- Yu Gu, Xiang Deng, and Yu Su. Don't generate, discriminate: A proposal for grounding language models to real-world environments, 2023. 9
- Yanjiang Guo, Yen-Jen Wang, Lihan Zha, Zheyuan Jiang, and Jianyu Chen. Doremi: Grounding language model by detecting and recovering from plan-execution misalignment, 2023. 1
- Shibo Hao, Yi Gu, Haodi Ma, Joshua Jiahua Hong, Zhen Wang, Daisy Zhe Wang, and Zhiting Hu. Reasoning with language model is planning with world model. *arXiv preprint arXiv:2305.14992*, 2023. 9
- Wenlong Huang, Pieter Abbeel, Deepak Pathak, and Igor Mordatch. Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato (eds.), *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9118–9147. PMLR, 2022a. URL <https://proceedings.mlr.press/v162/huang22a.html>. 1, 3, 6, 9, 15
- Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan Tompson, Igor Mordatch, Yevgen Chebotar, Pierre Sermanet, Noah Brown, Tomas Jackson, Linda Luu, Sergey Levine, Karol Hausman, and Brian Ichter. Inner monologue: Embodied reasoning through planning with language models, 2022b. 9
- Wenlong Huang, Fei Xia, Dhruv Shah, Danny Driess, Andy Zeng, Yao Lu, Pete Florence, Igor Mordatch, Sergey Levine, Karol Hausman, and Brian Ichter. Grounded decoding: Guiding text generation with grounded models for robot control, 2023. 9
- Yuqian Jiang, Shiqi Zhang, Piyush Khandelwal, and Peter Stone. Task planning in robotics: an empirical comparison of pddl-based and asp-based systems. *Cornell University - arXiv, Cornell University - arXiv*, Apr 2018. 9
- Leslie Pack Kaelbling and Tomás Lozano-Pérez. Hierarchical task and motion planning in the now. In *2011 IEEE International Conference on Robotics and Automation*, pp. 1470–1477, 2011. doi: 10.1109/ICRA.2011.5980391. 1, 3
- Angelos Katharopoulos, Apoorv Vyas, Nikolaos Pappas, and François Fleuret. Transformers are rnns: Fast autoregressive transformers with linear attention, 2020. 9
- J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977. ISSN 0006341X, 15410420. URL <http://www.jstor.org/stable/2529310>. 6
- Yaniv Leviathan, Matan Kalman, and Yossi Matias. Fast inference from transformers via speculative decoding, 2023. 9
- Shuang Li, Xavier Puig, Chris Paxton, Yilun Du, Clinton Wang, Linxi Fan, Tao Chen, De-An Huang, Ekin Akyürek, Anima Anandkumar, Jacob Andreas, Igor Mordatch, Antonio Torralba, and Yuke Zhu. Pre-trained language models for interactive decision-making. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022a. URL <https://openreview.net/forum?id=FWMQYjFso-a>. 3, 14
- Xiang Lorraine Li, Adhiguna Kuncoro, Jordan Hoffmann, Cyprien de Masson d'Autume, Phil Blunsom, and Aida Nematzadeh. A systematic investigation of commonsense knowledge in large language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 11838–11855, Abu Dhabi, United Arab Emirates, December 2022b. Association for Computational Linguistics. doi: 10.18653/v1/2022.emnlp-main.812. URL <https://aclanthology.org/2022.emnlp-main.812>. 3

- Jacky Liang, Wenlong Huang, Fei Xia, Peng Xu, Karol Hausman, Brian Ichter, Pete Florence, and Andy Zeng. Code as policies: Language model programs for embodied control, 2023. 9
- Bill Yuchen Lin, Yicheng Fu, Karina Yang, Prithviraj Ammanabrolu, Faeze Brahman, Shiyu Huang, Chandra Bhagavatula, Yejin Choi, and Xiang Ren. Swiftsage: A generative agent with fast and slow thinking for complex interactive tasks, 2023a. 9
- Bill Yuchen Lin, Chengsong Huang, Qian Liu, Wenda Gu, Sam Sommerer, and Xiang Ren. On grounded planning for embodied tasks with language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 13192–13200, 2023b. 9
- Lajanugen Logeswaran, Yao Fu, Moontae Lee, and Honglak Lee. Few-shot subgoal planning with language models. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 5493–5506, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.402. URL <https://aclanthology.org/2022.naacl-main.402>. 9
- Jieyi Long. Large language model guided tree-of-thought, 2023. 9
- Yao Mu, Qinglong Zhang, Mengkang Hu, Wenhai Wang, Mingyu Ding, Jun Jin, Bin Wang, Jifeng Dai, Yu Qiao, and Ping Luo. Embodiedgpt: Vision-language pre-training via embodied chain of thought, 2023. 9
- Vishal Pallagani, Bharath Muppasani, Keerthiram Murugesan, Francesca Rossi, Biplav Srivastava, Lior Horesh, Francesco Fabiano, and Andrea Loreggia. Understanding the capabilities of large language models for automated planning. *arXiv preprint arXiv:2305.16151*, 2023. 9
- Xavier Puig, Kevin Ra, Marko Boben, Jiaman Li, Tingwu Wang, Sanja Fidler, and Antonio Torralba. Virtualhome: Simulating household activities via programs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 3, 5, 10, 14
- Shreyas Sundara Raman, Vanya Cohen, Eric Rosen, Ifrah Idrees, David Paulius, and Stefanie Tellex. Planning with large language models via corrective re-prompting. *CoRR*, abs/2211.09935, 2022. doi: 10.48550/arXiv.2211.09935. URL <https://doi.org/10.48550/arXiv.2211.09935>. 1
- Noah Shinn, Beck Labash, and Ashwin Gopinath. Reflexion: an autonomous agent with dynamic memory and self-reflection. *arXiv preprint arXiv:2303.11366*, 2023. 1, 9
- Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. Progprompt: Generating situated robot task plans using large language models, 2022. 6, 9, 15
- Chan Hee Song, Jiaman Wu, Clayton Washington, Brian M. Sadler, Wei-Lun Chao, and Yu Su. Llm-planner: Few-shot grounded planning for embodied agents with large language models, 2023. 1, 9
- Haotian Sun, Yuchen Zhuang, Lingkai Kong, Bo Dai, and Chao Zhang. Adaplaner: Adaptive planning from feedback with language models. *arXiv preprint arXiv:2305.16653*, 2023a. 9
- Simeng Sun, Yang Liu, Shuohang Wang, Chenguang Zhu, and Mohit Iyyer. Pearl: Prompting large language models to plan and execute actions over long documents. *arXiv preprint arXiv:2305.14564*, 2023b. 9
- Mirac Suzgun, Luke Melas-Kyriazi, and Dan Jurafsky. Prompt-and-rerank: A method for zero-shot and few-shot arbitrary textual style transfer with small language models. In *arXiv*, 2022. 9
- Sai Vemprala, Rogerio Bonatti, Arthur Buckner, and Ashish Kapoor. Chatgpt for robotics: Design principles and model abilities. *Microsoft Autonomous Systems and Robotics Research*, 2023. 2, 6, 9
- Sinong Wang, Belinda Z. Li, Madian Khabsa, Han Fang, and Hao Ma. Linformer: Self-attention with linear complexity, 2020. 9

- Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *The Eleventh International Conference on Learning Representations*, 2023a. URL <https://openreview.net/forum?id=1PL1NIMMrw>. 9
- Zihao Wang, Shaofei Cai, Anji Liu, Xiaojian Ma, and Yitao Liang. Describe, explain, plan and select: Interactive planning with large language models enables open-world multi-task agents, 2023b. 9
- Yue Wu, So Yeon Min, Yonatan Bisk, Ruslan Salakhutdinov, Amos Azaria, Yuanzhi Li, Tom Mitchell, and Shrimai Prabhunoye. Plan, eliminate, and track – language models are good teachers for embodied agents, 2023a. 9
- Zhenyu Wu, Ziwei Wang, Xiuwei Xu, Jiwen Lu, and Haibin Yan. Embodied task planning with large language models, 2023b. 9
- Danfei Xu, Roberto Martín-Martín, De-An Huang, Yuke Zhu, Silvio Savarese, and Li Fei-Fei. Regression planning networks. *arXiv: Artificial Intelligence, arXiv: Artificial Intelligence*, Sep 2019. 1
- Sherry Yang, Ofir Nachum, Yilun Du, Jason Wei, Pieter Abbeel, and Dale Schuurmans. Foundation models for decision making: Problems, methods, and opportunities, 2023. 9
- Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022. 9
- Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models, 2023. 9
- Andy Zeng, Maria Attarian, brian ichter, Krzysztof Marcin Choromanski, Adrian Wong, Stefan Welker, Federico Tombari, Aveek Purohit, Michael S Ryoo, Vikas Sindhwani, Johnny Lee, Vincent Vanhoucke, and Pete Florence. Socratic models: Composing zero-shot multimodal reasoning with language. In *The Eleventh International Conference on Learning Representations*, 2023. URL <https://openreview.net/forum?id=G2Q2Mh3avow>. 9
- Zirui Zhao, Wee Sun Lee, and David Hsu. Large language models as commonsense knowledge for large-scale task planning, 2023. 9

# Appendix

## A ENVIRONMENT

### A.1 ACTION SPACE

The action types in the Virtual Home Environment are listed as follows:

- |            |             |           |               |
|------------|-------------|-----------|---------------|
| 1. Sleep   | 8. Wipe     | 15. Touch | 22. Drop      |
| 2. StandUp | 9. Pull     | 16. Open  | 23. Lie       |
| 3. WakeUp  | 10. Push    | 17. Close | 24. SwitchOn  |
| 4. Walk    | 11. Pour    | 18. Run   | 25. SwitchOff |
| 5. Find    | 12. TurnTo  | 19. Sit   | 26. Drink     |
| 6. Grab    | 13. PointAt | 20. Read  | 27. PutIn     |
| 7. Wash    | 14. Watch   | 21. PutOn | 28. PutBack   |

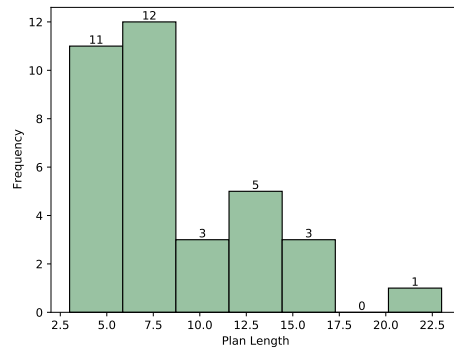


Figure 6: Distribution of plan length.

### A.2 PARTIAL OBSERVATION

We implemented partial observation based on [Li et al. \(2022a\)](#). The official definition of partial observation is that when the agent is situated in a room, its observation consists of all the objects in the room that are not located inside enclosed objects. For instance, if an apple is placed inside a closed refrigerator, it will not appear in the observation of the agent.

### A.3 OBSERVATION REPRESENTATION

In the VirtualHome Environment ([Puig et al., 2018](#)), observations primarily consist of two components: *object states* and *inter-object relationships*. The *object states* describes the state in which an object exists. For example, the state of television can either be “on” or “off”, i.e., “*On(TV)*” or “*Off(TV)*”. The *inter-object relationships* are represented using predicates to express the relationships between objects. For example, when a character is in close proximity to a television, there may be a predicate such as: “*Close(character, TV)*”. We convert the observations from VH into English sentences using a rule-based approach. For example, the predicate “*Close(character, TV)*” is transformed into “character is close to TV”, and the predicate “*Off(TV)*” is transformed into “TV is off”.

#### A.4 BASIC STATISTICS

In the gold plan annotated in the dataset, the plan with the longest length consists of 23 actions, while the average length is 8.13. The frequency distribution histogram regarding the length of the plan is shown in Figure 6. Furthermore, we have also computed the frequency histograms for actions and objects, which are depicted in Figure 7 and Figure 8.

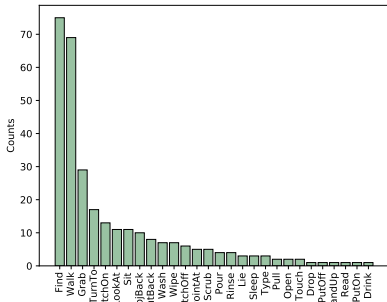


Figure 7: Histogram of action frequencies. Certain actions exhibit a significantly higher frequency than others, such as **Find** and **Walk**.

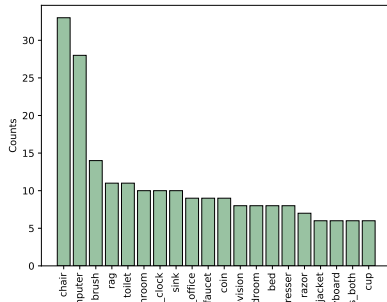


Figure 8: Histogram of object frequencies. Only objects with a frequency in the top 20 are included for better visualization.

### B MORE IMPLEMENTATION DETAILS

#### B.1 HYPER-PARAMETER SEARCH

For both Grounded Deciding and ITERATIVE-PLANNER, we conducted a grid search over the sampling parameters of OpenAI APIs. The search range for temperature was set from 0.1 to 1.0 in increments of 0.1, while the search range for top-p was set to 0.85, 0.9, 0.95, and 1.0.

In the case of Grounded Deciding, the optimal hyperparameter combination was found to be a temperature of 0.7 and topp of 1.0. As for ITERATIVE-PLANNER, the optimal hyperparameter combination was a temperature of 0 and topp of 1.0.

#### B.2 BASELINE MODELS

**ZERO-SHOT PLANNER** (Huang et al., 2022a) propose to translate each unstructured action generated by LLM into an admissible action via another pre-trained masked language model. The translated action is then appended to the prompt used for generating the remaining steps. We utilize the official implementation provided by Huang et al. (2022a), which employs a dynamically retrieved plan as an exemplar in the prompt. Moreover, concerning the hyper-parameters, we configure the maximum number of steps as 20 and set the early stopping threshold to 0.5 to achieve optimal performance.

**PROGPROMPT** (Singh et al., 2022) proposes a programming language-inspired prompt with an assert statement. These assert statements provide a mechanism for incorporating environment feedback into the plan generation process, ensuring that preconditions for each action are met;

#### B.3 POST-PROCESSING

To ensure the generated plan is actionable in the environment, we further implement a post-processing module after plan sampling: (i) **Format Checking:** Actions that do not conform to the required format and thus cannot be parsed, is discarded. Given the strong format following ability of GPT-3.5, the number of discarded items is minimal; (ii) **Object & Action Translation:** Even with the correct format, the plan generated by the LLM might include actions or objects not present in the environment. This issue often arises from semantically accurate but not exactly matching results. For instance, if the environment’s object name is "food\_egg", but the generated action includes

"egg", this discrepancy requires resolution. Firstly, we parse the LLM’s action string to identify the action and object names. Then we use BERT similarity to match these with the environment’s available actions and objects <sup>3</sup>. For example, for the LLM’s generated string '[Switch On] <telly >', the action parser identifies "Switch On" as the action and "telly" as the object. These are then matched with available actions and objects to result in the actionable action "[SwitchOn] <TV >".

## C MORE EXPERIMENTAL RESULTS

### C.1 RESULTS BY PLAN LENGTH

	EXEC. $\uparrow$	SR $\uparrow$	GCR $\uparrow$	No.EC $\downarrow$
$0 <  a  \leq 5$	100.00 $\pm$ 0.00	64.72 $\pm$ 5.20	77.12 $\pm$ 4.13	0.30 $\pm$ 0.15
$5 <  a  \leq 10$	83.59 $\pm$ 4.03	35.10 $\pm$ 5.95	52.25 $\pm$ 3.33	2.47 $\pm$ 0.35
$10 <  a  \leq 15$	88.09 $\pm$ 8.48	26.98 $\pm$ 4.89	55.98 $\pm$ 7.00	3.20 $\pm$ 1.12
$15 <  a $	66.67 $\pm$ 0.00	22.22 $\pm$ 0.00	36.11 $\pm$ 0.00	4.09 $\pm$ 0.21

Table 4: The performance of TREE-PLANNER<sub>N=50</sub> across different plan sequence lengths.

Table 4 above presents the performance of TREE-PLANNER<sub>N=50</sub>, categorized by different plan sequence lengths. In general, as the plan sequence length increases, the performance of the model tends to decrease. Specifically, for tasks with plan lengths smaller than 5, the SR can reach 64.72% and even 100% for EXEC.. However, for tasks with plan lengths larger than 15, the SR decreases to 22.22% (-42.5%), and EXEC. decreases to 66.67% (-33.33%). Furthermore, as the plan length increases, the number of corrections in the model also increases. This is due to the higher likelihood of errors accumulating in longer sequential task planning, thus necessitating a greater need for error correction. The above experimental results provide a more comprehensive analysis of the performance of our approach, emphasizing potential directions for future enhancements. Furthermore, we have provided the results of LOCAL REPLAN as shown in Table 5. It can be observed that while TREE-PLANNER maintains comparable performance across different plan lengths, it exhibits a significant advantage in terms of the number of corrections.

	EXEC. $\uparrow$	SR $\uparrow$	GCR $\uparrow$	No.EC $\downarrow$
$0 <  a  \leq 5$	94.17 $\pm$ 2.04	56.94 $\pm$ 0.79	69.46 $\pm$ 1.70	1.05 $\pm$ 0.32
$5 <  a  \leq 10$	77.02 $\pm$ 7.63	38.10 $\pm$ 3.57	49.66 $\pm$ 5.45	3.92 $\pm$ 0.56
$10 <  a  \leq 15$	69.84 $\pm$ 2.97	24.78 $\pm$ 5.83	44.52 $\pm$ 2.16	5.14 $\pm$ 0.17
$15 <  a $	72.22 $\pm$ 28.33	22.22 $\pm$ 0.00	35.65 $\pm$ 8.36	4.83 $\pm$ 0.87

Table 5: The performance of LOCAL REPLAN across different plan sequence lengths.

### C.2 RESULTS BY SCENE

As is shown in Table 6, TREE-PLANNER exhibits consistent performance across various scenes, thereby further illustrating its robustness.

	EXEC. $\uparrow$	SR $\uparrow$	GCR $\uparrow$	No.EC $\downarrow$
ENV-1	92.42 $\pm$ 2.14	45.45 $\pm$ 3.71	64.72 $\pm$ 3.25	1.74 $\pm$ 0.44
ENV-2	91.30 $\pm$ 4.10	36.75 $\pm$ 4.10	52.43 $\pm$ 7.13	2.33 $\pm$ 0.47
ENV-3	85.18 $\pm$ 2.62	37.04 $\pm$ 2.62	50.80 $\pm$ 3.19	1.83 $\pm$ 0.39
ENV-4	88.89 $\pm$ 3.14	48.89 $\pm$ 3.14	66.74 $\pm$ 1.72	2.35 $\pm$ 0.58

Table 6: The performance of TREE-PLANNER<sub>N=50</sub> across different scenes.

<sup>3</sup><https://www.sbert.net/>



### C.3 DIVERSITY OF SAMPLED PLANS

As shown in Figure 9, the number of different plans varies approximately linearly with the change in the sampling  $n$ . Therefore, when conducting plan sampling, the issue of homogenization due to the sampled plans will not arise.

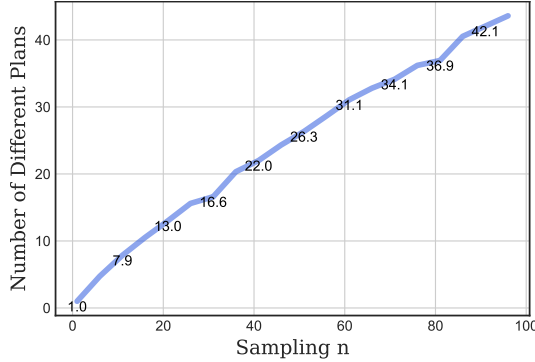


Figure 9: The diversity of plans generated. The x-axis represents the number of programs sampled, while the y-axis represents the average number of different plans generated.

### C.4 UPPER LIMIT OF PLAN SAMPLING

We also compute the corresponding Success Rate (SR) result of Figure 5 by: (i) the maximum SR for all generated plans, i.e.,  $SR_{max}(c) = \max_{i=1}^N(SR(c_i))$ ; (ii) the average SR for all generated plans, i.e.,  $SR_{avg}(c) = \frac{1}{N} \sum_{i=1}^N(SR(c_i))$ . As is shown in Figure 10, the trend of SR concerning  $N$  closely aligns with the trajectory of GCR.

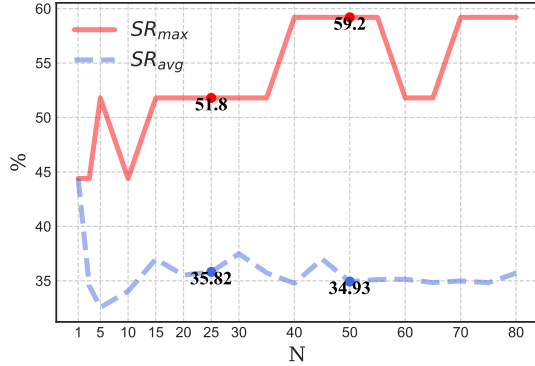


Figure 10: Maximum and average Success Rate (SR) for all sampled plans.

### C.5 EXPERIMENTAL RESULTS WITH BEST-FIRST SEARCH

**Language-Guided Best-First Search** The edges of the action tree are assigned weights based on the log probabilities of the actions they lead to. These weights are aggregated when multiple paths converge into the same action, thereby reflecting the cumulative likelihood of reaching that action within the context of the tree’s structure. The evaluation function of Language-Guided Best-First Search is defined as:  $f_L(a_i) = \frac{e^{\log(\text{Prob}(a_i))}}{\sum_j e^{\log(\text{Prob}(a_j))}}$ . Here,  $a_i$  represents one of the child nodes of a given node and iterates over all child nodes. The softmax function normalizes the log probabilities, making them suitable for probabilistic comparison and selection.

**Environment-Guided Best-First Search** Actions are weighted based on the observability of their associated objects. If all objects involved in an action are observable in the environment, the action

is assigned a higher weight (1), otherwise, a lower weight (0). To facilitate comparative analysis and to account for varying numbers of actions and objects, a softmax function is applied to the weights of actions emanating from each node. The evaluation function of Environment-Guided Best-First Search is defined as  $f_E(a_i) = \frac{e^{\text{Observability}(a_i)}}{\sum_j e^{\text{Observability}(a_j)}}$ .

**Hybrid Best-First Search** The Hybrid Method combines the probabilistic language and environment observability factors. The evaluation function is:  $f_{\text{Hybrid}}(a_i) = \alpha f_L(a_i) + (1 - \alpha) f_E(a_i)$  In this function,  $\alpha$  is a hyper-parameter between 0 and 1 that balances the influence of the language-based probabilities and the environmental observability

**Experimental Setup** We conducted experiments in the case of  $N=50$  without error correction. As for the hyper-parameter  $\alpha$ , we heuristically set it to 0.5.

**Experimental Results** As demonstrated by Table 7, although heuristic methods exhibit a greater advantage in terms of token efficiency, their SR falls short compared to grounded deciding (using LLM as an implicit evaluation function). Moreover, combining both evaluation functions yields even greater performance improvements, thus demonstrating that there is still room for enhancement through the optimization of the evaluation function in heuristic methods.

Table 7: Experimental Results with Best-First Search

	EXEC. $\uparrow$	SR $\uparrow$	GCR $\uparrow$	\$COST $\downarrow$
TREE-PLANNER	49.01 $\pm$ 5.67	28.14 $\pm$ 2.45	35.84 $\pm$ 4.20	3.48 $\pm$ 0.04
LANGUAGEGUIDED	34.71 $\pm$ 0.65	15.06 $\pm$ 1.02	22.47 $\pm$ 1.72	2.16 $\pm$ 0.02
ENVIRONMENTGUIDED	37.26 $\pm$ 0.17	6.24 $\pm$ 0.65	12.10 $\pm$ 1.41	2.16 $\pm$ 0.02
HYBRID $_{\alpha=0.5}$	38.71 $\pm$ 0.42	17.74 $\pm$ 0.69	25.34 $\pm$ 0.95	2.16 $\pm$ 0.02

## C.6 PROPORTION OF NEW OBJECTS

The relationship between the number of steps and the proportion of new objects in evaluated tasks is shown in Figure 11.

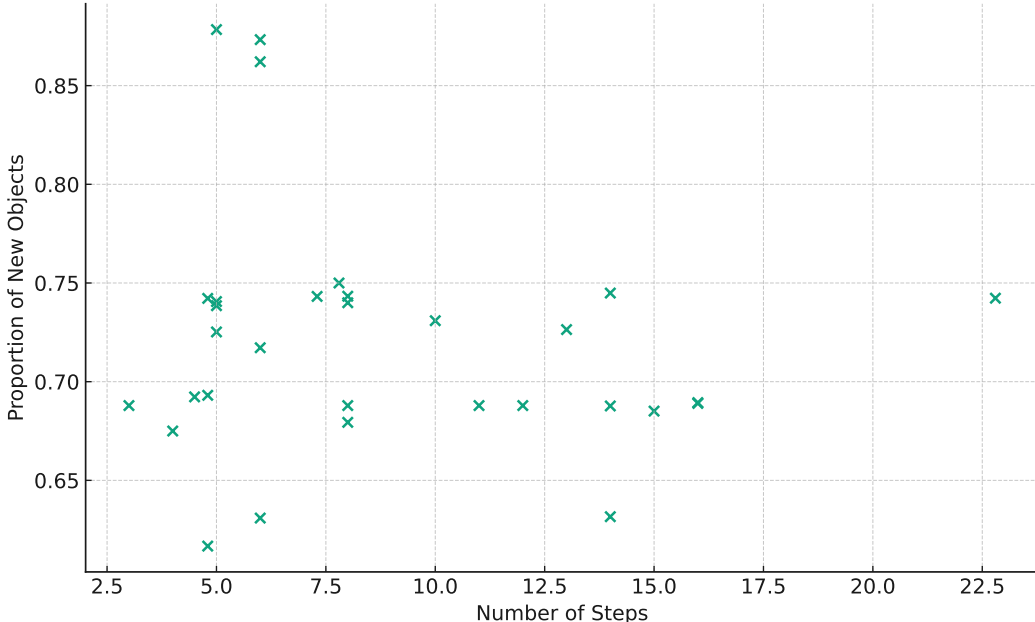


Figure 11: Each point represents a distinct task. The proportion of new objects is calculated with gold plans in the dataset.

## D DETAILS ON TOKEN EFFICIENCY

The detailed derivation of the boundary conditions for  $N$  is as follows:

$$\begin{aligned}
 & \text{to prove } \mu_{ours} < \mu_{sbs} \\
 \Rightarrow & \rho_{ps} + MN|a| + M \cdot (\rho_{gd} + N) < M \cdot (\rho_{ps} + \rho_{gd} + |a|) \\
 \Rightarrow & MN|a| + M \cdot N < (M - 1) \cdot \rho_{ps} + M \cdot |a| \\
 \Rightarrow & M \cdot (|a| + 1) \cdot N < (M - 1) \cdot \rho_{ps} + M \cdot |a| \\
 \Rightarrow & N < \frac{1 - 1/M}{1 + 1/|a|} \cdot \frac{\rho_{ps}}{|a|} + \frac{|a|}{|a| + 1}
 \end{aligned}$$

## E QUALITATIVE ANALYSIS ON ERRORS

### E.1 INCORRECT DECIDING

**Task:** *Hang up jacket*

**Action Tree:** See Figure 12

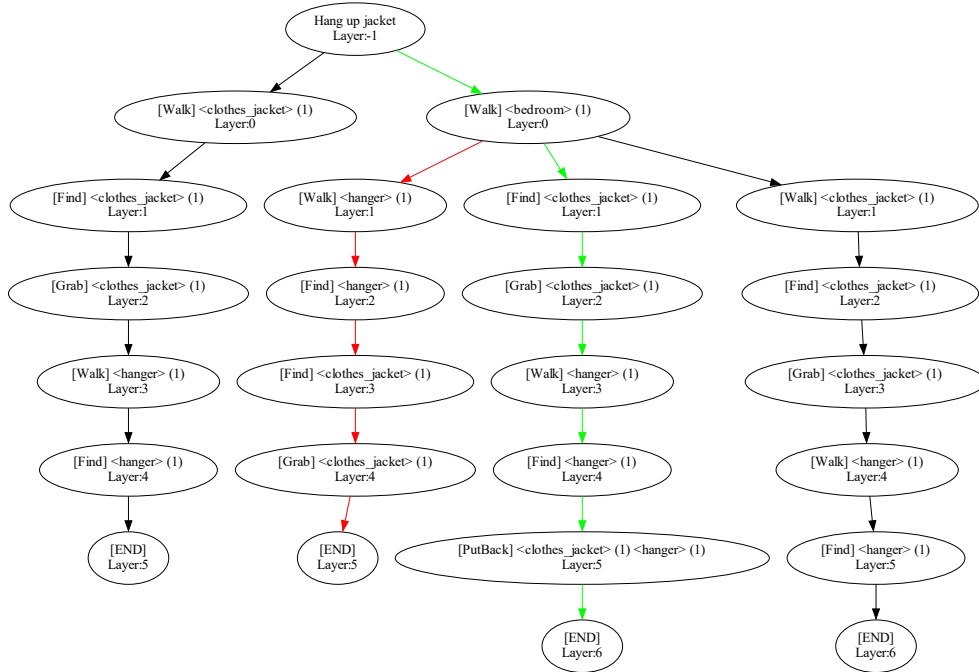


Figure 12: Visualization of the action tree for *Hang up jacket*. The red path is the plan chosen by LLM, while the green path is the correct plan.

**Explanation:** As depicted in Figure 12, the model made an incorrect decision during the second step. This is due to the failure of TREE-PLANNER to comprehend the sequential order of searching for a jacket and searching for a hanger.

### E.2 ENVIRONMENT MISUNDERSTANDING

**Task:** *Clean toilet*

**Example Plan:**

- |                         |                           |
|-------------------------|---------------------------|
| 1. [Walk] <bathroom>(1) | 7. [Wash] <toilet>(1)     |
| 2. [Walk] <toilet>(1)   | 8. [Find] <mop_bucket>(1) |
| 3. [Pull] <toilet>(1)   | 9. [Walk] <detergent>(1)  |
| 4. [Wash] <toilet>(1)   | 10. [Grab] <detergent>(1) |
| 5. [Wipe] <toilet>(1)   | 11. [Pour] <detergent>(1) |
| 6. [Push] <toilet>(1)   | <mop_bucket>(1)           |

**Explanation:** The toilet is not movable, so the execution of step 3 would be failed.

## E.3 INCOMPLETE PLAN

**Task:** *Put alarm clock in bedroom*

**Example Plan:**

1. [Walk] <bedroom>(1)
2. [Walk] <alarm\_clock>(1)
3. [Find] <alarm\_clock>(1)
4. [Grab] <alarm\_clock>(1)
5. [Find] <dresser>(1)
6. [Open] <dresser>(1)
7. [SwitchOn] <alarm\_clock>(1)

**Explanation:** After executing this plan, the alarm clock will be in the hand of the character. However, a step is missing: [PutBack] the <alarm\_clock>(1) on the <dresser>(1) to place the alarm clock *in bedroom*.

## E.4 ILLOGICAL ERROR

**Task:** *Shave*

**Example Plan:**

- |                              |                                |
|------------------------------|--------------------------------|
| 1. [Walk] <bathroom>(1)      | 7. [Walk] <razor>(1)           |
| 2. [Find] <shaving_cream>(1) | 8. [Find] <razor>(1)           |
| 3. [Grab] <shaving_cream>(1) | 9. [Grab] <razor>(1)           |
| 4. [Walk] <after_shave>(1)   | 10. [PutOn] <shaving_cream>(1) |
| 5. [Find] <after_shave>(1)   | 11. [Wash] <headset>(1)        |
| 6. [Grab] <after_shave>(1)   | 12. [PutOn] <after_shave>(1)   |

**Explanation:** The character does not have a free hand when executing step 9. This is because, in steps 3 and 6, the character has already picked up two items.

## E.5 SEMANTICALLY CORRECT

**Task:** *Use computer*

**Example Plan:**

1. [Walk] <home\_office>(1)
2. [Walk] <computer>(1)

3. **[Find]** <chair>(1)
4. **[Sit]** <chair>(1)
5. **[SwitchOn]** <computer>(1)
6. **[Type]** <keyboard>(1)

**Explanation:** The error information from the environment is as follows: “<keyboard >” cannot be found when executing “[Type] <keyboard >(1)”. However, at this moment, the character is seated in front of the computer, indicating that the keyboard should be located near the character.

## E.6 MORE RESULTS ON ERROR TYPES

As illustrated in Table 8, we conducted a further analysis of the error types in the TREE-PLANNER<sub>N=25</sub> model *with correction*. The distribution is strikingly similar to that of the model *without correction*. Moreover, owing to the presence of error correction, the model demonstrates enhanced capability in avoiding *semantically correct* errors during the grounded deciding stage.

## F PROMPTS

### F.1 ITERATIVE-PLANNER

**(Instruction)** You need to act as a task planner who decomposes a high-level household task into several sub-tasks. The temporal relationship between subtask sequences must adhere to common-sense logic. Each sub-task can be one of the following form: 1. [action\_name]; 2. [action\_name] <object name 1>(object id 1). 3. [action\_name] <object name 1>(object id 1) <object name 2>(object id 2). The number of arguments depends on the action type. The (object id) is used to tell the simulator that the actions should be done on the same object instance. For example a program as: [Walk] <glass>(1) [Grab] <glass>(1) Indicates that the agent should first walk to a glass and then grab that same glass. If you think your task has been successful, you can output [END], which is action type 1.

**(Global Information)** For action type 1, the available actions are: [Sleep], [StandUp], [WakeUp] For action type 2, the available actions are: [Walk], [Find], [Grab], [Wash], [Wipe], [Pull], [Push], [Pour], [TurnTo], [PointAt], [Watch], [Touch], [Open], [Close], [Run], [Sit], [Read], [PutOn], [Drop], [Lie], [SwitchOn], [SwitchOff], [Drink] For action type 3, the available actions are: [PutIn], [PutBack] All action\_name of the sub-tasks must be chosen from the above actions, and follow the corresponding format. You are in a house that consists of four rooms. These rooms are bathroom, dining\_room, bedroom, home\_office. Available objects in the house are : clothes\_hat, ceilinglamp, cpuscreen, orchid, couch, trashcan, dresser, dishwasher, centerpiece, phone, toaster, measuring\_cup, stereo, mat, computer, envelope, oven\_mitts, piano\_bench, box, photoframe, shower, ceiling, wall, window, freezer, faucet, detergent, light, desk, napkin, food\_rice, kitchen\_counter, folder, stovefan, walllamp, food\_food, coffee\_pot, food\_steak, jelly, vacuum\_cleaner, powersocket, filing\_cabinet, alcohol, bathroom, door, bathroom\_counter, clothes\_gloves, microwave, oven, sink, milk, ice, bedroom, laptop, doorjamb, food\_cake, bills, tea\_bag, television, laser\_pointer, toilet, board\_game, sponge, food\_carrot, table, tray, cupboard, mousepad, picture, tvstand, tablelamp, hanger, pot,

Table 8: Distribution of error types of the TREE-PLANNER<sub>N=25</sub> *with correction* model.

Error Type	Explanation	Proportion(%)
Missing Correct Plans	Plan sampling did not yield correct plans	<b>52.6%</b>
Environment Misunderstanding	Misunderstandings on actions or objects	21.1%
Incomplete Plan	The absence of essential steps	15.8%
Illogical Error	The generated plan is logically incorrect	5.3%
Semantically Correct	Execution failed but semantically correct	10.5%
Grounded Deciding Error	Errors during grounded deciding	<b>47.4%</b>
Incorrect Deciding	Incorrect decisions at specific nodes	47.4%
Semantically Correct	Execution failed but semantically correct	0.0%

dry\_pasta, floor, knifeblock, curtain, chair, food\_bread, drawing, creditcard, check, coffe\_maker, character, pasta, bag, food\_bacon, bookshelf, toothbrush\_holder, cutting\_board, home\_office, dining\_room, nail\_polish, pillow, tape, nightstand, bathroom\_cabinet, bench, conditioner, cat, bed, keyboard, mouse All object names must be chosen from the above object list

**(Observation)** Currently, you are standing in the bedroom, and holding nothing in your right hand and nothing in your left hand. pillow is clean. napkin is clean. pillow is dirty. bed is clean. mat is dirty. pillow is close to drawing. tablelamp is close to bed. mat is facing drawing. pillow is inside bedroom. mat is close to table. bed is close to drawing. table is close to mat. pillow is on floor. floor is close to bed. tablelamp is close to pillow. mat is close to curtain. bed is facing computer. mat is close to floor. pillow is facing drawing. curtain is close to mat. bed is close to tablelamp. wall is close to mat. napkin is inside bedroom. window is close to mat. pillow is close to tablelamp. bed is close to floor. pillow is close to wall. mat is close to filing\_cabinet. drawing is close to bed. pillow is close to floor. bed is close to wall. filing\_cabinet is close to mat. bed is close to nightstand. mat is inside bedroom. pillow is close to pillow. pillow is close to nightstand. mat is close to wall. wall is close to pillow. nightstand is close to pillow. nightstand is close to bed. drawing is close to pillow. floor is close to pillow. bed is inside bedroom. floor is close to mat. wall is close to bed. mat is close to window. pillow,napkin,pillow,mat,bed,pillow,pillow is inside bedroom.

**(In-Context Examples)**

Task: Watch TV

[Find] <remote\_control>(1)  
[Find] <television>(1)  
[SwitchOn] <television>(1)  
[Find] <couch>(1)  
[Sit] <couch>(1)  
[Touch] <remote\_control>(1)  
[TurnTo] <television>(1)  
[LookAt] <television>(1)

Task: Turn on light

[Walk] <dining\_room>(1)  
[Walk] <light>(1)  
[Find] <light>(1)  
[SwitchOn] <light>(1)  
[Find] <light>(2)  
[SwitchOn] <light>(2)

Task: Go to sleep

[Walk] <bedroom>(1)  
[Walk] <bed>(1)  
[Lie] <bed>(1)  
[Sleep]

Task: Brush teeth

[Walk] <bathroom>(1)  
[Walk] <toothbrush\_holder>(1)  
[Find] <toothbrush\_holder>(1)  
[Find] <toothbrush>(1)  
[Grab] <toothbrush>(1)  
[Walk] <tooth\_paste>(1)  
[Find] <tooth\_paste>(1)  
[Grab] <tooth\_paste>(1)  
[Pour] <tooth\_paste>(1) <toothbrush>(1)  
[Find] <teeth>(1)  
[Scrub] <teeth>(1)

Task: Take nap  
 [Walk] <bedroom>(1)

## F.2 PLAN SAMPLING

**(Instruction)** You need to act as a task planner, who decompose a high-level household task into several sub-tasks. The temporal relationship between subtask sequences must adhere to common-sense logic. Each sub-task can be one of the following form: 1. [action\_name]; 2. [action\_name] <object name 1 >(object id 1). 3. [action\_name] <object name 1 >(object id 1) <object name 2 >(object id 2). The number of arguments depends on the action type. The (object id) is used to tell the simulator that the actions should be done on the same object instance. For example a program as: [Walk] <glass>(1) [Grab] <glass>(1) Indicates that the agent should first walk to a glass, and then grab that same glass.

**(Global Information)** For action type 1, the available actions are: [Sleep], [StandUp], [WakeUp] For action type 2, the available actions are: [Walk], [Find], [Grab], [Wash], [Wipe], [Pull], [Push], [Pour], [TurnTo], [PointAt], [Watch], [Touch], [Open], [Close], [Run], [Sit], [Read], [PutOn], [Drop], [Lie], [SwitchOn], [SwitchOff], [Drink] For action type 3, the available actions are: [PutIn], [PutBack] All action\_name of the sub-tasks must be chosen from the above actions, and follow the corresponding format. You are in a house that consists of four rooms. These rooms are bathroom, dining\_room, bedroom, home\_office. Available objects in the house are : clothes\_hat, ceilinglamp, cpuscreen, orchid, couch, trashcan, dresser, dishwasher, centerpiece, phone, toaster, measuring\_cup, stereo, mat, computer, envelope, oven\_mitts, piano\_bench, box, photoframe, shower, ceiling, wall, window, freezer, faucet, detergent, light, desk, napkin, food\_rice, kitchen\_counter, folder, stovefan, walllamp, food\_food, coffee\_pot, food\_steak, jelly, vacuum\_cleaner, powersocket, filing\_cabinet, alcohol, bathroom, door, bathroom\_counter, clothes\_gloves, microwave, oven, sink, milk, ice, bedroom, laptop, doorjamb, food\_cake, bills, tea\_bag, television, laser\_pointer, toilet, board\_game, sponge, food\_carrot, table, tray, cupboard, mousepad, picture, tvstand, tablelamp, hanger, pot, dry\_pasta, floor, knifeblock, curtain, chair, food\_bread, drawing, creditcard, check, coffe\_maker, character, pasta, bag, food\_bacon, bookshelf, toothbrush\_holder, cutting\_board, home\_office, dining\_room, nail\_polish, pillow, tape, nightstand, bathroom\_cabinet, bench, conditioner, cat, bed, keyboard, mouse All object names must be chosen from the above object list

**(Initial Observation)** Currently, you are standing in the home\_office, and holding nothing in your right hand and nothing in your left hand.

### **(In-Context Examples)**

Task: Watch TV  
 [Find] <remote\_control>(1)  
 [Find] <television>(1)  
 [SwitchOn] <television>(1)  
 [Find] <couch>(1)  
 [Sit] <couch>(1)  
 [Touch] <remote\_control>(1)  
 [TurnTo] <television>(1)  
 [LookAt] <television>(1)

Task: Turn on light  
 [Walk] <dining\_room>(1)  
 [Walk] <light>(1)  
 [Find] <light>(1)  
 [SwitchOn] <light>(1)  
 [Find] <light>(2)  
 [SwitchOn] <light>(2)

Task: Go to sleep  
 [Walk] <bedroom>(1)  
 [Walk] <bed>(1)  
 [Lie] <bed>(1)

**[Sleep]**

Task: Brush teeth

**[Walk]** <bathroom>(1)**[Walk]** <toothbrush\_holder>(1)**[Find]** <toothbrush\_holder>(1)**[Find]** <toothbrush>(1)**[Grab]** <toothbrush>(1)**[Walk]** <tooth\_paste>(1)**[Find]** <tooth\_paste>(1)**[Grab]** <tooth\_paste>(1)**[Pour]** <tooth\_paste>(1) <toothbrush>(1)**[Find]** <teeth>(1)**[Scrub]** <teeth>(1)

Task: Take nap

**F.3 GROUNDED DECIDING**

**(Instruction)** You need to act as a home robot. At each moment, I will provide you with observations of your current environment, as well as the high-level task I want you to do, and previous mid-level sub-tasks that have been executed. Then, you need to select the best sub-task from the options I provide to complete the designated home task based on the observation and your past experience. When one choosed sub-task causes an error in the environment, you will be provided with the error information and the corresponding sub-task, and you need to re-choose a corrective sub-task at the current time step. For example, The sub-tasks that have been executed in the environment are:

**[GRAB]** <plate>(1)**[WALK]** <dining room>(1)The choosed sub-task is: **[PUTBACK]** <plate>(1) <table>(1)

The prompt (error information) would be: The sub-task: **”[PUTBACK] <plate>(1) <table>(1)”** caused an error: Script is not executable, since <character>(1) is not close to <table>(1) when executing **”[PUTBACK] <plate>(1) <table>(1) [1]”** Among the following actions, which action would you take.

A. **[Find]** <table>(1)B. **[Find]** <plate>(1)

A corrective choice of sub-task would be (You just need to provide the mark before the option you want to choose): A

**(Observation)** Currently, you are standing in the bedroom, and holding nothing in your right hand and nothing in your left hand. pillow is clean. napkin is clean. pillow is dirty. bed is clean. mat is dirty. pillow is close to drawing. tablelamp is close to bed. mat is facing drawing. pillow is inside bedroom. mat is close to table. bed is close to drawing. table is close to mat. pillow is on floor. floor is close to bed. tablelamp is close to pillow. mat is close to curtain. bed is facing computer. mat is close to floor. pillow is facing drawing. curtain is close to mat. bed is close to tablelamp. wall is close to mat. napkin is inside bedroom. window is close to mat. pillow is close to tablelamp. bed is close to floor. pillow is close to wall. mat is close to filing\_cabinet. drawing is close to bed. pillow is close to floor. bed is close to wall. filing\_cabinet is close to mat. bed is close to nightstand. mat is inside bedroom. pillow is close to pillow. pillow is close to nightstand. mat is close to wall. wall is close to pillow. nightstand is close to pillow. nightstand is close to bed. drawing is close to pillow. floor is close to pillow. bed is inside bedroom. floor is close to mat. wall is close to bed. mat is close to window. pillow,napkin,pillow,mat,bed,pillow,pillow is inside bedroom.

Your task is: Take nap.

**(History)**

Your previously executed sub-tasks are:

**[Walk]** <bedroom>(1)

Among the following sub-tasks, which one would you take.

A. **[FIND]** <bed>(1)



- B. [WALK] <couch>(1)  
 C. [WALK] <bed>(1)  
 D. [FIND] <couch>(1)  
 E. [FIND] <pillow>(1)  
 The best choice of sub-task is:

## G VISUALIZED ACTION TREE

We visualized the action trees for the tasks listed in Table 9 for interested readers.

Task	Action Tree
<i>Take nap</i>	See Figure 13
<i>Put on your shoes</i>	See Figure 14
<i>Pick up spare change on dresser</i>	See Figure 15
<i>Clean toilet</i>	See Figure 16
<i>Put alarm clock in bedroom</i>	See Figure 17

Table 9: Action trees for various tasks.

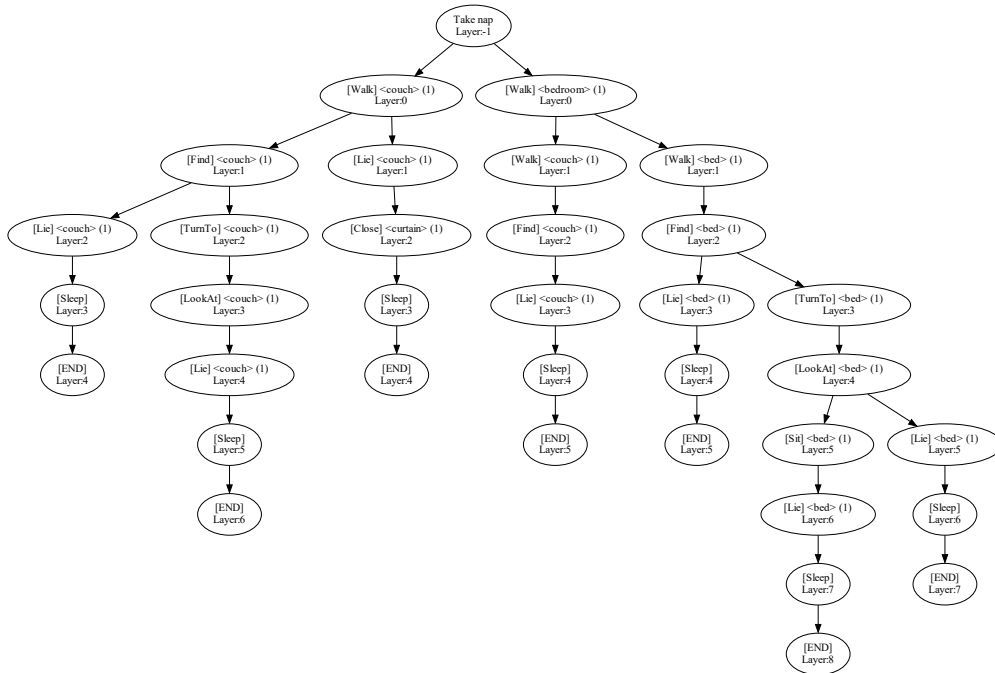
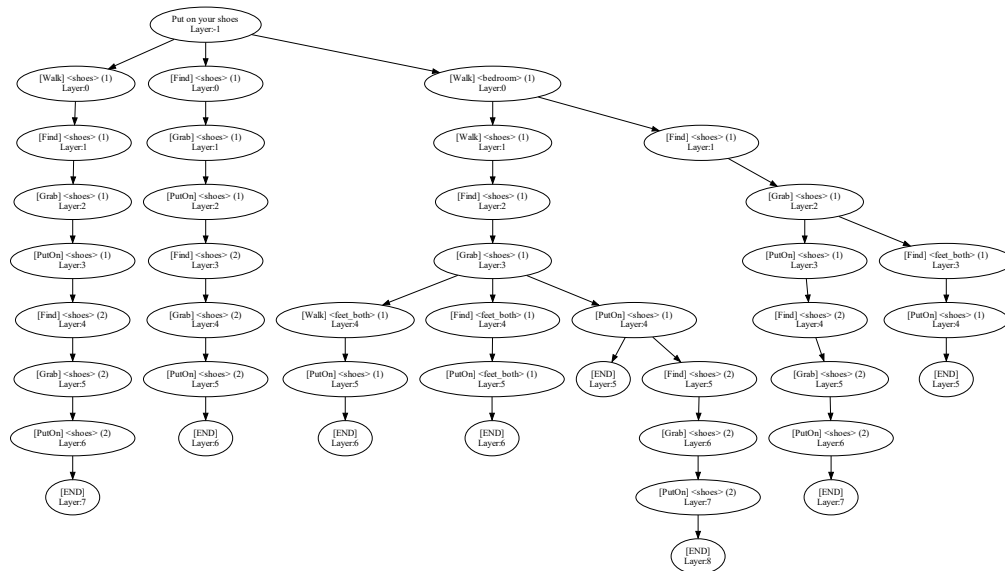


Figure 13: Visualization of the action tree for *Take nap*.

Figure 14: Visualization of the action tree for *Put on your shoes*.

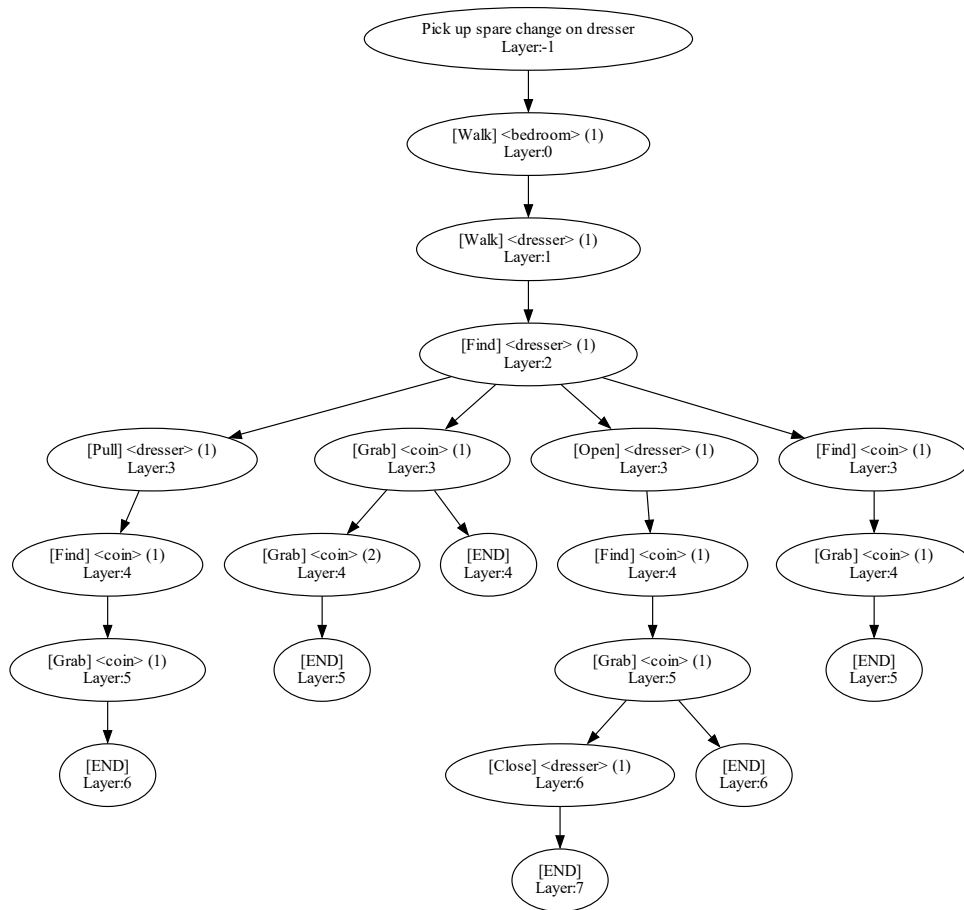


Figure 15: Visualization of the action tree for *Pick up spare change on dresser*.

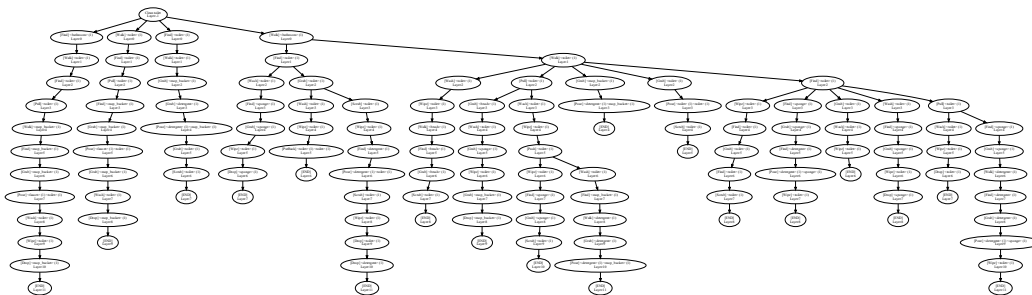


Figure 16: Visualization of the action tree for *Clean toilet*.

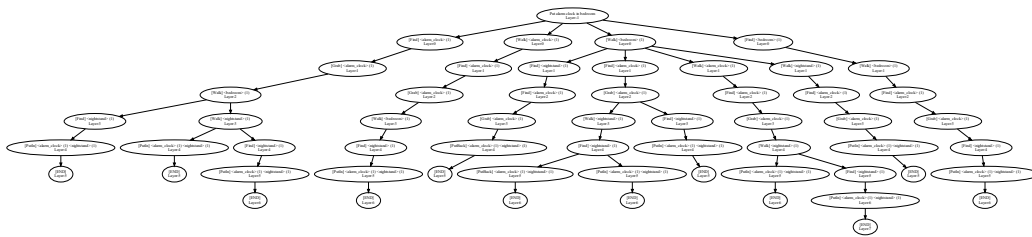


Figure 17: Visualization of the action tree for *Put alarm clock in bedroom*.