

000 001 002 003 004 005 006 007 008 009 010 011 012 013 014 015 016 017 018 019 020 021 022 023 024 025 026 027 028 029 030 LEARNING TO PLAN WITH PERSONALIZED PREFERENCES

005 **Anonymous authors**

006 Paper under double-blind review

009 ABSTRACT

012 Effective integration of Artificial Intelligence (AI) agents into daily life requires
 013 them to understand and adapt to individual human preferences, particularly in as-
 014 sistive roles. Although recent studies on embodied intelligence have advanced sig-
 015 nificantly, they typically adopt generalized approaches that overlook personalized
 016 preferences in planning. Cognitive research has demonstrated that these prefer-
 017 ences serve as crucial intermediate representations in human decision-making pro-
 018 cesses and, though implicitly expressed through minimal demonstrations, can gen-
 019 eralize across diverse planning scenarios. To systematically address this gap, we
 020 introduce the Preference-based Planning (PBP) benchmark, an embodied bench-
 021 mark designed to evaluate agents’ ability to learn preferences from few demon-
 022 strations and adapt their planning strategies accordingly. PBP features hundreds
 023 of diverse preferences spanning from atomic actions to complex sequences, en-
 024 abling comprehensive assessment of preference learning capabilities. Evaluations
 025 of SOTA methods reveal that while symbol-based approaches show promise in
 026 scalability, significant challenges remain in learning to generate plans that sat-
 027 isfy personalized preferences. Building on these findings, we develop agents that
 028 not only learn preferences from few demonstrations but also adapt their planning
 029 strategies based on these preferences. Experiments in PBP demonstrate that in-
 030 corporating learned preferences as intermediate representations significantly im-
 proves an agent’s ability to construct personalized plans, establishing preference
 as a valuable abstraction layer for adaptive planning.

031 032 033 1 INTRODUCTION

034 The field of embodied Artificial Intelligence (AI) is rapidly advancing, driven by significant progress
 035 in foundation models for vision and language (Bommasani et al., 2021; Peng et al., 2023; Achiam
 036 et al., 2023; Bai et al., 2023). These advances enable AI systems to autonomously assist or collab-
 037 orate with humans in daily tasks, particularly in domestic settings (Driess et al., 2023; Leal et al.,
 038 2023; Zitkovich et al., 2023; Ahn et al., 2024). However, recent approaches utilizing natural lan-
 039 guage instructions (Mu et al., 2023; Zitkovich et al., 2023; Singh et al., 2023) face fundamental lim-
 040 itations in capturing human preferences (Zhu et al., 2016). While natural language is our primary
 041 means of communication, its inherent ambiguity creates a gap between instructions and intended
 042 executions (Yuan et al., 2022; Jiang et al., 2022; 2021; Yuan et al., 2020). For instance, when a user
 043 requests help in preparing an apple, the agent needs to understand specific preferences about apple
 044 selection, washing requirements, cutting style, and container choice—details that vary significantly
 045 across individuals; see also Figure 1 for a graphical illustration.

046 Preference, central to personalization (Slovic, 1995), remains inadequately addressed in embodied
 047 Artificial Intelligence (AI). Integrating personalized preferences is crucial for tailoring agent actions
 048 to individual users, thereby enhancing the effectiveness and satisfaction of embodied assistants (Lee
 049 et al., 2012; Leyzberg et al., 2014). Moreover, preferences guide human-like decision-making and
 050 intelligent behavior. Psychological research emphasizes that understanding preferences is vital for
 051 interpreting human behaviors (Fawcett & Markson, 2010) and facilitating social interactions (Gerson
 052 et al., 2017; Liberman et al., 2021), suggesting that preference understanding could enable more
 053 grounded planning in embodied assistants.

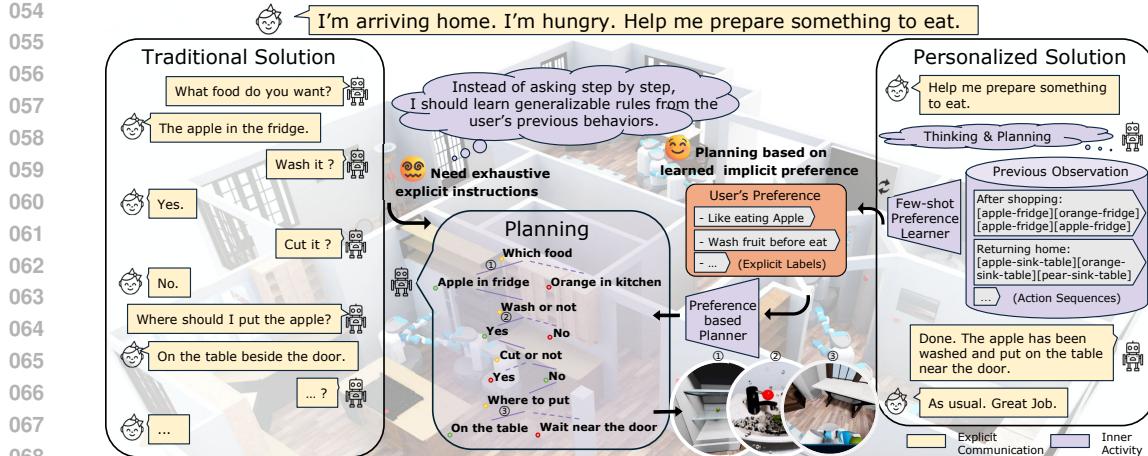


Figure 1: **An example of preference-based planning in a food preparation scenario.** When the assistant receives a natural language instruction for a food preparation task, it can follow one of two approaches: (Left, traditional methods) The assistant verifies details with the user at each step through exhaustive communication; or (Right, our personalized approach) it first learns from previous user action sequences to infer explicit preference labels and then generates a personalized plan based on the learned preferences. The planning tree (middle) illustrates how preferences guide the whole decision-making process across multiple dimensions. By learning preferences as a key intermediate representation from minimal human demonstrations, our approach enables AI agents to deliver personalized and adaptable assistance without explicit step-by-step instructions.

Learning human preferences in real-world settings presents unique challenges (Peng et al., 2024). Humans typically communicate their needs succinctly (Levinson, 1983), without exhaustive preference details (Lichtenstein & Slovic, 2006), and many preferences include unconscious or instinctive elements difficult to articulate (Epstein, 1994; Simonson, 2008). A practical approach is to infer preferences from observed human choices and decision-making patterns, as illustrated in Figure 1, where a robot assistant can learn users' preferences and behavior habits from previous observations.

In this paper, we focus on agents capable of learning preferences from human behavior and subsequently planning actions guided by these learned preferences. While previous studies have explored preference-based learning, they are limited to specific tasks or non-embodied scenarios (e.g., rearrangement (Kapelyukh & Johns, 2022), scheduling (Yuan et al., 2023), dialogue generation (Ashby et al., 2023)) and fail to generalize across different situations. To address this limitation, we introduce Preference-based Planning (PBP), a comprehensive embodied benchmark built upon NVIDIA Omniverse and OmniGibson (Li et al., 2023a). PBP provides realistic simulation and real-time rendering for thousands of daily activities across 50 scenes, featuring a parameterized vocabulary of 290 diverse preferences. These preferences span multiple levels, from specific action-level preferences (e.g., preferred glass type, water temperature) to task sequence-level preferences (e.g., task ordering, subtask prioritization).

Given the expensive nature of data collection (Akgun et al., 2012) and the few-shot nature of preference acquisition, we frame preference learning as a few-shot learning from demonstration task. In this framework, agents must respond to ambiguous instructions by formulating plans aligned with preferences demonstrated in limited example sequences. Specifically, an agent needs to analyze behavioral data, identify consistent patterns, and extrapolate these patterns to higher-level preference abstractions that can generalize across various tasks (Chao et al., 2011). Furthermore, when confronted with new tasks, the agent should leverage these learned preferences to generate adaptive action sequences that align with user preferences while maintaining task efficiency.

With the PBP benchmark developed, we challenge existing learning agents on their ability to learn human preference and subsequently conduct preference-based planning. Our systematic evaluation of State-of-the-Art (SOTA) algorithms on PBP reveals that preferences serve as valuable abstractions of human behaviors, and their incorporation as intermediate planning steps significantly enhances agent adaptability. Through extensive experimentation, we demonstrate that symbol-based approaches show promise in scalability, yet significant challenges remain in both preference learning and planning. These challenges stem from the complexity of planning intricate activities and the nuanced nature of learning preferences through perception. Our analysis particularly highlights the

108 difficulties in few-shot preference learning and preference-guided planning, establishing preferences
 109 as a crucial abstraction layer between high-level goals and low-level actions. We present this work
 110 as a foundation for addressing these challenges in preference-based embodied AI.

112 2 RELATED WORK

114 2.1 THEORETICAL FOUNDATIONS OF HUMAN PREFERENCES

116 Preference theory originates from psychological research, where it describes predictable patterns
 117 in human behavior that can be modeled mathematically (Kahneman, 1982). These preferences re-
 118 flect individual attitudes towards available choices in decision-making (Lichtenstein & Slovic, 2006)
 119 and operate both consciously and unconsciously to shape behavior (Coppin et al., 2010). A funda-
 120 mental principle is that underlying preferences can be inferred from consistent behavioral patterns
 121 (Sen, 1973), enabling systematic analysis of decision-making processes. This framework has ex-
 122 tended beyond psychology into economics, where Rational Choice Theory (Scott et al., 2000) mod-
 123 els decision-making based on rational self-interest (Zey, 1998). Building on this, Utility Theory
 124 provides a mathematical foundation for modeling how preferences relate to attitudes toward rewards
 125 and risks (Mongin, 1997; Aleskerov et al., 2007). These theoretical foundations establish prefer-
 126 ences as fundamental elements in shaping both individual behavior and broader societal dynamics.
 127 In recent years, these preference models have found new applications in artificial intelligence and
 128 robotics, particularly in developing human-centric AI assistants capable of understanding and adapt-
 129 ing to individual user preferences.

130 2.2 EMBODIED TASK PLANNING BENCHMARKS

132 The development of intelligent embodied planning has evolved from basic Vision-and-Language
 133 Navigation (VLN) tasks (Anderson et al., 2018; Chen et al., 2019; Thomason et al., 2020) to com-
 134 plex interactive scenarios. ALFRED (Shridhar et al., 2020) introduced object manipulation, state
 135 tracking, and temporal dependencies between instructions, and platforms like Habitat (Savva et al.,
 136 2019; Puig et al., 2023b) and AI2-THOR (Kolve et al., 2017) emphasize active perception, long-
 137 term planning, and interactive learning, while VisualAgentBench (Liu et al., 2024b) and Embodied-
 138 Bench Yang et al. (2025) focus on abilities of visual foundation agents. Recent research has also
 139 shifted toward implicit-instruction scenarios particularly in arrangement (Taniguchi et al., 2021;
 140 Kant et al., 2022; Sarch et al., 2022), where agents reason about object placements without ex-
 141 plicit directives. These works leverage commonsense knowledge or **general preferences** to execute
 142 universally-accepted behavioral norms, but neglect nuanced variations among different people.

143 Our work emphasizes **personalized preferences**, where embodied agents align their actions with
 144 individual user habits. This includes personalized object placement strategies (Abdo et al., 2015;
 145 Kapelyukh & Johns, 2022; Wu et al., 2023) in housekeeping, multi-agent coordination where agents
 146 adapt to individual preferences to achieve optimal coordination (Shu & Tian, 2019; Puig et al.,
 147 2021b; 2023a). Works on proactive assistance (Patel & Chernova, 2023; Patel et al., 2023) further
 148 explore anticipating temporal patterns in humans' daily routines. We extend these chanllenges by
 149 considering preferences across diverse situations and scenes. Beyond object placements, we address
 150 temporal action sequences, state transitions during interactions, and few-shot preference learning,
 151 thus further enabling robust preference modeling and adaptation in real-world scenarios.

152 2.3 PERSONALIZATION IN FOUNDATION MODELS

153 The problem of adapting foundation models to distinct users has raised great attention (Tseng et al.,
 154 2024). Existing studies explore prompting methods or Reinforcement Learning from Human Feed-
 155 back (RLHF) to align models to user preferences in recommendation (Yang et al., 2023; Wang &
 156 Lim, 2023), education (Huber et al., 2024; Park et al., 2024), healthcare (Zhang et al., 2024b), and
 157 dialogue generation (Li et al., 2023b; Ashby et al., 2023; Song et al., 2020; Lee et al., 2024). However,
 158 while these foundation models excel at learning and adapting to individual preferences in conversa-
 159 tional settings, their ability in the situation of limited demonstrations and embodied planning remains
 160 an open challenge, particularly in multi-step tasks requiring complex perceptual understanding and
 161 subtle execution strategies. Utilizing Large Language Models (LLMs) as few-shot planners to gen-
 162 erate language-based action sequences from limited demonstrations (Song et al., 2023; Driess et al.,

162 2023; Ding et al., 2023; Zhang et al., 2024a), or using Vision-Language Models (VLMs) to enhance
 163 robotic systems' perception and reasoning capabilities (Ahn et al., 2024; Leal et al., 2023; Gu et al.,
 164 2023; Brohan et al., 2022; Zitkovich et al., 2023; Xu et al., 2024a) has proven to be effective. Yet,
 165 how to incorporate personalization into them for embodied planning is not fully addressed.
 166

167 3 FORMULATING PREFERENCE-BASED PLANNING

170 Tasks in PBP mirror real-world watch-and-help scenarios (Puig et al., 2021a), where an agent ob-
 171 serves a few demonstrations of a user performing tasks that reveal preferences. The agent must then
 172 complete similar tasks in different setups while adhering to the demonstrated preferences.

173 Preference-based planning comprises two key components: few-shot **preference learning** of user
 174 preferences and subsequent **planning** guided by these learned preferences. Since humans, even
 175 infants, can naturally detect others' preferences from limited decisions (Choi & Luo, 2023), and
 176 collecting extensive personal demonstrations is impractical in daily life, we formulate this as few-
 177 shot learning from demonstration.

178 Given a user with preference \mathbf{p} , the agent observes the user performing tasks from a first-person
 179 perspective, denoted as \mathbf{O} . These observations span multiple demonstrations. Formally, \mathbf{O} contains
 180 both state and action observations: $\mathbf{O} = \{(\mathcal{S}_i, \mathcal{A}_i, \mathcal{M})_N\}$, where \mathcal{S}_i denotes the egocentric ob-
 181 servation sequence in the i -th demonstration, \mathcal{A}_i represents the action sequence, and \mathcal{M} optionally
 182 provides a bird's-eye view of the entire scene map.

183 In the first stage, the objective is to learn the preference representation demonstrated through user
 184 actions:

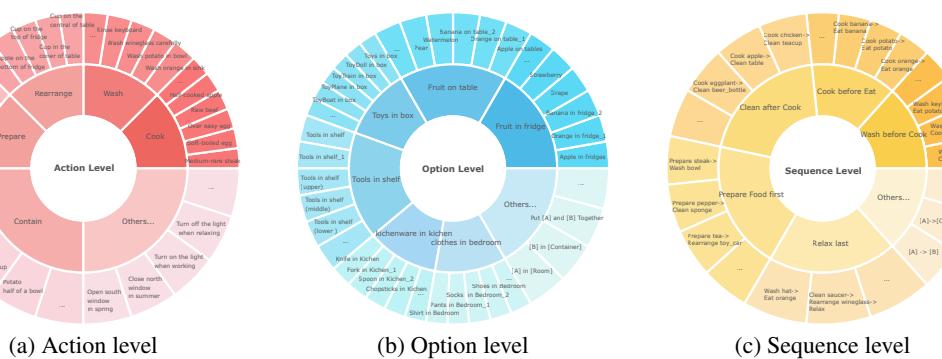
$$\mathbf{p} = f(\mathbf{O}; \theta_f), \quad (1)$$

187 where \mathbf{p} denotes the learned preference representation here. It can either be a hidden representation
 188 or an explicit textual label, depending on the task settings.

189 The learned preference \mathbf{p} should then guide planning when the agent faces different setups with
 190 varying objects, room layouts, or entire scenes. Specifically, the agent optimizes:

$$\mathcal{L} = \sum_{i=1} \ell(g(s_i, f(\mathbf{O}; \theta_f); \theta_g), a_i), \quad (2)$$

194 where $g(\cdot)$ represents a potentially parameterized planning function that maps the current state and
 195 preference representation to the next action, and a_i denotes the ground-truth action demonstrating
 196 the user's preference at the current stage.



209 **Figure 2: Hierarchical organization of user preferences.** Our framework organizes preferences in a
 210 three-tiered structure, visualized through sunburst diagrams: (a) Action level captures fine-grained execution details
 211 within specific tasks, from quantity preferences in "Contain" (e.g., "half a cup" vs. "full cup") to environmental
 212 controls (e.g., lighting and window operations). (b) Option level represents spatial preferences for object
 213 categories, encoding both storage decisions (e.g., table vs. fridge for fruits) and organizational choices (e.g.,
 214 shelf levels and boxes for tools/toys). (c) Sequence level defines temporal relationships between tasks, encom-
 215 passing both basic preparation sequences (e.g., "Prepare Food first") and conditional orderings (e.g., "Clean
 216 after Cook," "[A]->[B]"). Each diagram's hierarchical structure branches from general categories to specific
 217 instances, revealing detailed preference patterns upon closer inspection. (Vector graphics; zoom in for details.)

216 4 THE PREFERENCE-BASED PLANNING (PBP) BENCHMARK

218 Built on NVIDIA’s Omniverse and OmniGibson simulation environment (Li et al., 2023a), our PBP
 219 benchmark enables realistic simulation of thousands of daily activities. It spans 50 distinct scenes
 220 and encodes 290 unique preferences, with a comprehensive test set of 5000 instances. Below, we
 221 detail the preference structure and test set construction.

223 4.1 DEFINITION OF PREFERENCES

225 We organize preferences in a three-tiered hierarchical structure that captures varying degrees of
 226 specificity across tasks. Figure 2 provides an overview of all preferences and their distribution,
 227 while Figure 3 illustrates concrete examples of preferences and corresponding agent actions. The
 228 290 preferences are distributed across three levels: 80 for sequence-level, 135 for option-level, and
 229 75 for action-level preferences.

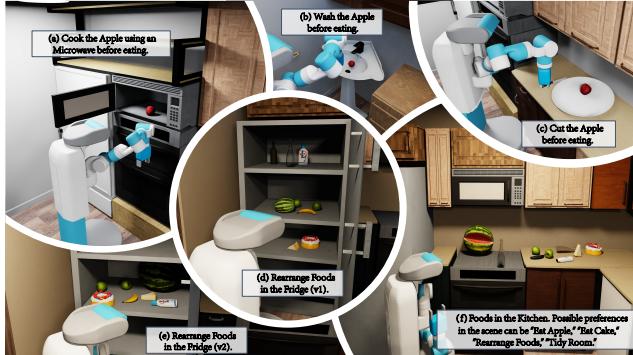
230 **Action Level** These bottom-level preferences govern fine-grained execution details within specific
 231 sub-tasks, such as water quantity preferences when filling cups or shelf placement choices for books.

232 **Option Level** Middle-level preferences encode alternative approaches to sub-tasks. For instance,
 233 in “storing-nonperishable-food,” users may prefer cabinet storage versus table placement. These
 234 preferences can bind to different objects and may compose multiple action-level preferences.

235 **Sequence Level** Top-level preferences define task ordering and prioritization. They capture tempo-
 236 ral dependencies between sub-tasks, such as cleaning furniture before rearranging kitchen utensils,
 237 followed by dinner preparation upon returning home.

239 4.2 CONSTRUCTING PBP TEST SET

241 Our PBP benchmark includes a de-
 242 fault test set for systematic evalua-
 243 tion. Following the formulation
 244 in Section 3, we structure
 245 PBP tasks as few-shot learning-
 246 from-demonstration problems. Each
 247 test point comprises several (typi-
 248 cally three) unique demonstrations
 249 with egocentric observations of ac-
 250 tion sequences and their corre-
 251 sponding preference labels. As illus-
 252 trated in Figure 4, a demonstration includes
 253 an egocentric video of agent activi-
 254 ty, a bird’s-eye-view map tracking
 255 agent position, and frame-level action
 256 annotations. We also provide third-
 257 person view recordings for enhanced
 258 visualization. We prioritize the ego-
 259 centric perspective for two reasons:
 260 1) it offers a clear view with minimal
 261 occlusions, and 2) it aligns with human perception, facilitating transfer to real-world data from
 262 head-mounted devices.



263 **Figure 3: Example of preferences and their corresponding actions in PBP.** Demonstration of preference hierarchies in PBP:
 264 (a-c) Primitive actions (cooking, washing, cutting); (d-e) Object
 265 rearrangement strategies (category grouping vs. fridge-layer place-
 266 ment); (f) Task sequencing (fruits before cleaning). (Vector graph-
 267 ics; zoom in for details.)

268 The test set construction follows a two-stage process. First, we build a reusable and extensible
 269 demonstration pool. To generate each demonstration, we randomly assign a preference from our
 270 defined primitives to one of 50 OmniGibson scenes, then sample relevant objects within the chosen
 271 scene. We generate multi-perspective observations using rule-based planners for high-level planning
 272 and predefined scripts for low-level execution (e.g., Inverse Kinematics (IK) for grasping, A* for
 273 movement).

274 Second, we construct test points by sampling preferences and retrieving relevant demonstrations
 275 from the pool. To reflect real-world few-shot scenarios, each preference is paired with demon-
 276 strations that share the same high-level preference but vary in scene settings or object selections. We
 277 also include unrelated demonstrations to prevent sampling bias.

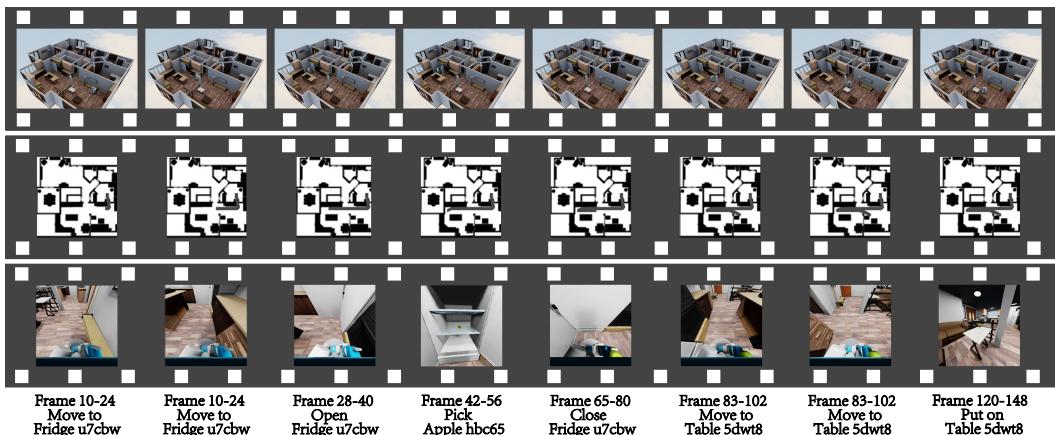


Figure 4: **Example of a demonstration in PBP.** The robot in the demonstration is executing the task “Pick Apple from Fridge and place on Table”. **Top:** A third-person view video provides an overhead perspective of the entire scene. **Middle:** The bird’s-eye-view map displays the robot’s relative position within the scene. **Bottom:** The egocentric video captures the robot’s first-person observations during task execution. **Text:** The per-frame action annotations contain Omniverse object IDs, which ensure each object reference is unique and enable the model to identify specific objects precisely.

The default test set contains 5,000 test points, drawing from a pool of 15,000 unique recordings. Unless specified otherwise, all experiments use this default set. The benchmark also supports custom test point generation through flexible demonstration sampling, preference definition, and third-person view video creation.

4.3 MODELS

Our evaluation focuses primarily on multimodal models that incorporate LLMs and demonstrate strong few-shot learning capabilities. The LLM component serves as a knowledge base that can enhance preference learning through commonsense reasoning. We also include symbol-based LLM models for ablation studies to analyze how different modalities impact PBP performance. Most models evaluated can function in both end-to-end and two-stage pipeline configurations. See Appendix D for detailed implementations.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETUP

We evaluate preference learning capabilities across two distinct settings: end-to-end and two-stage approaches. In the end-to-end setting, models directly map raw state inputs to action outputs. Leveraging models’ in-context learning abilities, we provide demonstrations alongside current state information as input and evaluate the generated action sequences against ground truth.

The two-stage setting introduces an intermediate step where models first learn to predict explicit preference labels during training. These predicted labels then serve as preference representations for subsequent planning stages. For black-box models, we employ carefully designed prompts rather than fine-tuning approaches.

All demonstration videos maintain consistent technical specifications across models and agents: egocentric perspective, 512×512 resolution, and 8 fps frame rate. Video duration matches the corresponding action sequence length. For LLM inference, we use conservative decoding parameters: temperature of 0.05, top-k of 1, and top-p of 0.05. All experiments run on a single machine with 8 NVIDIA A100 GPUs.

5.2 END-TO-END ACTION PREFERENCE LEARNING

We first evaluate model performance in the end-to-end setting, where models generate actions directly from previous demonstrations and current state information. To quantify performance, we

324
 325 **Table 1: Levenshtein distance between generated and ground truth action sequences.** **End-to-end** repre-
 326 sents models directly generating action sequences from demonstration-preference pair examples. **Two-stage**
 327 indicates generation using both demonstrations and previously inferred preference labels based on demonstra-
 328 tions. **Second-stage (gt)** uses demonstrations alongside ground truth preference labels for sequence generation.

	VIDEO-BASED INPUT				SYMBOL-BASED INPUT		
	ViViT	LLaVA-Next	EILEV	GPT-4V	Llama3-8B	DeepSeek-R1	GPT-4.1
Option Level							
End-to-end	15.49 \pm 1.29	15.94 \pm 3.41	12.88 \pm 2.20	15.63 \pm 2.31	14.74 \pm 3.21	8.73 \pm 3.03	7.42 \pm 2.67
Two-stage	-	12.46 \pm 3.23	12.89 \pm 3.74	8.37 \pm 2.19	9.67 \pm 5.16	3.19 \pm 2.19	2.26 \pm 2.03
Second-stage (gt)	-	3.28 \pm 5.29	11.18 \pm 4.20	1.26 \pm 2.55	8.22 \pm 5.58	1.76 \pm 1.89	0.15 \pm 2.85
Sequence Level							
End-to-end	34.04 \pm 11.84	34.76 \pm 11.25	33.10 \pm 12.21	33.75 \pm 11.15	31.79 \pm 7.32	28.72 \pm 4.12	26.48 \pm 3.25
Two-stage	-	30.02 \pm 13.54	33.03 \pm 13.61	27.52 \pm 9.48	25.46 \pm 5.93	18.61 \pm 3.25	14.19 \pm 3.01
Second-stage (gt)	-	18.92 \pm 14.18	26.57 \pm 12.21	11.36 \pm 8.05	19.02 \pm 7.10	14.10 \pm 3.76	10.31 \pm 2.98
Overall							
End-to-end	24.76	25.35	22.99	24.69	23.26	18.72	16.95
Two-stage	-	21.24	22.96	17.94	17.56	10.90	8.22
Second-stage (gt)	-	11.10	18.88	6.31	13.62	7.93	5.23

340 use Levenshtein distance to measure discrepancies between generated and ground truth action se-
 341 quences, treating each individual action as a token.

343 As shown in Table 1 (the **End-to-end** row), video-based models produce Levenshtein distances
 344 approaching the average ground truth sequence lengths (15.80 at option level, 35.87 at sequence
 345 level). These high distances indicate that the models generate predominantly inconsistent action
 346 sequences, suggesting a failure to grasp preferences embedded in demonstration videos. While
 347 symbol-based models show modest improvements, their performance gains remain limited.

348 These findings expose a fundamental limitation in current models: they struggle to extract underly-
 349 ing relationships from perceptual inputs without explicit intermediate guidance. The models appear
 350 to learn individual, isolated actions rather than cohesive action patterns that reflect implicit prefer-
 351 ences. This significant gap underscores the inherent challenge of performing end-to-end preference
 352 learning solely from demonstrations.

353 5.3 TWO-STAGE LEARNING-PLANNING

355 Given the limitations of end-to-end learning, we implement a two-stage approach to decompose the
 356 preference learning problem. The first stage focuses on preference prediction, where we provide
 357 models with auxiliary preference token labels and train them to predict hidden preferences explic-
 358 itely. These preference tokens, as discussed in Section 4.1, maintain sufficient semantic content for
 359 translation into primitive actions.

360 Results from the first stage (Table 2) reveal significant performance variations across models. At the
 361 option level, GPT-4V achieves superior performance with 48.48% accuracy, demonstrating strong
 362 capability in interpreting demonstrated preferences. Among symbol-based models, the stark contrast
 363 between DAG-Opt’s limited performance and the improved results from Llama3-8B and GPT-4.1
 364 highlights the advantage of next-token prediction over dependency learning for preference inference.
 365 Models with language components consistently show improved preference understanding compared
 366 to end-to-end learning.

367 **Table 2: Preference prediction accuracy in few-shot and ablative settings.**

	VIDEO-BASED INPUT				SYMBOL-BASED INPUT			
	ViViT	LLaVA-Next	EILEV	GPT-4V	DAG-Opt	Llama3-8B	DeepSeek-R1	GPT-4.1
Few-shot								
Option Level	9.38	36.87	38.33	48.48	10.15	72.98	86.02	88.91
Sequence Level	4.24	24.85	32.69	37.50	13.49	67.18	71.21	70.28
Overall	6.81	30.86	35.51	42.99	11.82	70.08	78.62	79.60
Ablative								
Option Level	9.16	15.47	4.77	29.42	3.84	39.50	78.19	75.29
Sequence Level	4.38	8.13	0.00	0.00	1.28	6.25	15.29	14.11
Overall	6.77	11.8	2.38	14.71	2.56	22.88	46.74	44.70

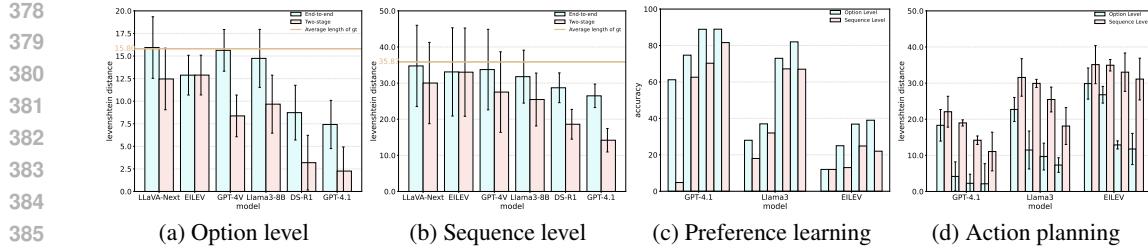


Figure 5: **(a)-(b) Levenshtein distance between generated and ground truth action sequences.** Results shown for both (a) option level and (b) sequence level under two conditions: End-to-end bars represent direct sequence generation from previous observations, while Two-stage bars show performance when models receive predicted preference labels. The — line indicates average ground truth sequence length (option level: 15.80, sequence level: 35.87). Results demonstrate significantly improved performance under the two-stage approach compared to end-to-end generation. **(c)-(d) Ablation study on the number of demonstrations.** Models are evaluated across both of the two stages: (c) first-stage preference learning and (d) second-stage action planning. We evaluate both Option Level and Sequence Level tasks. The number of few-shot demonstrations varies from [1, 2, 3, 5], presented left to right. For (c), higher accuracy indicates better performance. For (d), lower distance indicates better performance. Increased demonstration quantity generally improves both preference learning capability and planning effectiveness.

The second stage involves generating action sequences based on both demonstrations and predicted preference labels from the first stage, introducing potential error propagation. Results in Table 1 (Two-stage row) and Figure 5 (a)-(b) show significant improvements when models receive explicit preferences. For comprehensive evaluation, we include planning results using ground truth preference labels (Second-stage (gt) row). GPT-4V and GPT-4.1 achieve near-zero Levenshtein distances, indicating almost perfect alignment with ground truth action sequences.

Analysis of both stages reveals distinct challenges across model types. Vision-based models like LLaVA-Next and GPT-4V struggle with preference inference but excel in action planning given preference labels, suggesting difficulty in abstracting preferences from visual input. Symbol-based models perform well in both preference inference and preference-guided planning, yet underperform in end-to-end settings. This indicates that models may lack innate preference-based reasoning capabilities but can effectively plan when preferences are explicitly provided.

To isolate the impact of prior knowledge versus in-context learning, we conduct ablation studies by removing demonstrations and testing preference prediction on isolated test sequences. Results in Table 2 (bottom) show significant performance degradation compared to few-shot learning (Table 2 (top)), particularly at the sequence level. This suggests that while models may encode basic task-specific preferences, they rely heavily on demonstrations to recognize complex preference patterns in varied sequences.

Table 3: **Models’ generalization ability.** *direct* denotes experiments *without* generalization. *gen* denotes experiments conducted *with* generalization cases. We report the preference prediction accuracy.

	LLaVA-Next	EILEV	GPT-4V	DeepSeek-R1	GPT-4.1
Option Level <i>direct</i>	33.25	46.93	53.24	86.02	88.91
Option Level <i>gen</i>	36.87	38.33	48.48	84.98	87.12
Sequence Level <i>direct</i>	33.12	37.53	39.42	71.21	70.28
Sequence Level <i>gen</i>	24.85	32.69	37.50	70.16	68.01

5.4 GENERALIZATION

While human actions may vary across different objects and scenes, underlying preferences often remain consistent. We evaluate the models’ ability to generalize preference learning across varying visual contexts. The original test set inherently tests generalization by randomly sampling scenes and objects when rendering video demonstrations for each preference. To gain additional insights, we conduct complementary experiments with controlled conditions where demonstration and test videos are rendered in identical rooms with the same objects. This controlled setting enables direct performance comparisons under consistent conditions. We evaluate Efficient In-context Learning on Egocentric Videos (EILEV), Large Language and Vision Assistant (LLaVA), and GPT-4.1 series

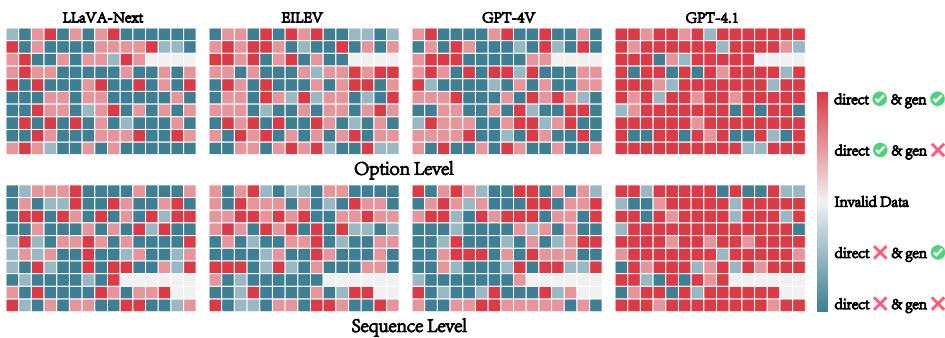


Figure 6: **Analysis of test samples in *direct* and *generalization* settings.** Lines represent distinct scenes, with grid colors indicating different sample statuses (*direct* for no generalization and *gen* for the opposite).

models on this variant of PBP, as these models previously demonstrated strong few-shot reasoning capabilities. Results are summarized in Table 3.

Symbol-based reasoning models demonstrate consistent performance regardless of scene or object variations, while vision-based models show greater sensitivity to scene changes. This distinction stems from the nature of our predefined preferences, which are sufficiently abstract and general to apply across diverse scenes and objects. Vision-based models, however, tend to anchor their few-shot learned preferences to specific visual features of scenes or objects. When these visual elements change, preference recognition accuracy may deteriorate. This contextual dependence remains a persistent challenge for vision-based models, which often overfit to scene-specific features from training videos.

Analysis of test points across *direct* and *gen* conditions (Figure 6) reveals two key findings: (i) Preference learning performance correlates with scene characteristics, with certain scenes proving consistently challenging across both conditions. (ii) While *direct* cases show better performance overall, failure patterns differ between conditions, particularly for vision-based models. This suggests models rely heavily on visual context consistency—including object arrangement and scene layout—for accurate predictions, indicating potential superficial learning rather than true preference understanding. Symbol-based reasoning maintains robust performance across varied scenes due to the general nature of predefined preferences, whereas vision-based models' strong dependence on specific visual contexts limits their generalization capability.

5.5 ABLATIONS ON DEMONSTRATION NUMBERS

We examine the effect of demonstration quantity on model performance through an ablation study (Figure 5 (c)-(d)). Results show that increasing demonstration numbers generally improves preference learning and planning effectiveness. This improvement is most evident in second-stage planning, where models achieve lower sequence distances by more accurately replicating human actions. Models like GPT-4.1, Llama3, and EILEV show consistent performance gains with additional demonstrations. However, we observe that excessive demonstrations (*e.g.*, 5-demo cases for GPT-4.1 and EILVE) can sometimes impair first-stage prediction accuracy. Despite these occasional exceptions, the overall trend confirms our intuition: more demonstrations enhance learning and planning performance. These findings highlight the importance of demonstration quantity in developing effective personalized planning systems that align with user preferences.

6 CONCLUSION

We investigate embodied agents that learn and implement human preferences through behavioral observation and user interaction. We present Preference-based Planning (PBP), a comprehensive embodied benchmark designed to capture the complexity of real-world human preferences. An evaluation framework is developed to assess models' preference learning and implementation capabilities. Our findings demonstrate that preferences effectively abstract human behaviors and guide planning processes. While current models still face challenges in preference inference and adaptive planning from limited observations, incorporating preference-based reasoning improves both effectiveness and generalization. We aim to stimulate further research in this crucial yet understudied domain of developing preference-aware embodied agents.

486
487

ETHICS STATEMENT

488
489

490 The development of personalized, preference-based embodied AI agents has the potential to significantly enhance human-AI collaboration, improving efficiency and user satisfaction in a variety of daily tasks. As AI agents increasingly adapt to individual behaviors, ensuring that these systems respect user autonomy and data privacy is crucial. Given our focus on private scenarios, we anticipate minimal negative societal impact from this research.

492
493
494

REPRODUCIBILITY STATEMENT

495
496
497

498 We have provided source code and running tutorials as supplementary materials. Key implementation details and experimental settings are described in the main paper (Section 5 and Appendix D). We will open-source both the code and the dataset used in our experiments upon acceptance.

498

499

REFERENCES

500
501
502

503 Nichola Abdo, Cyril Stachniss, Luciano Spinello, and Wolfram Burgard. Robot, organize my shelves! tidying
504 up objects by predicting user preferences. In *International Conference on Robotics and Automation (ICRA)*,
505 2015.

506
507
508

509 Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo
510 Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint*
511 *arXiv:2303.08774*, 2023.

512
513
514

515 Michael Ahn, Debidatta Dwibedi, Chelsea Finn, Montse Gonzalez Arenas, Keerthana Gopalakrishnan, Karol
516 Hausman, Brian Ichter, Alex Irpan, Nikhil Joshi, Ryan Julian, et al. Autorf: Embodied foundation models
517 for large scale orchestration of robotic agents. *arXiv preprint arXiv:2401.12963*, 2024.

518
519

520 Baris Akgun, Maya Cakmak, Karl Jiang, and Andrea L Thomaz. Keyframe-based learning from demonstration:
521 Method and evaluation. *International Journal of Social Robotics*, 4:343–355, 2012.

522
523

524 Fuad Aleskerov, Denis Bouyssou, and Bernard Monjardet. *Utility maximization, choice and preference*, volume 16. Springer Science & Business Media, 2007.

525
526
527

528 Peter Anderson, Qi Wu, Damien Teney, Jake Bruce, Mark Johnson, Niko Sünderhauf, Ian Reid, Stephen Gould,
529 and Anton Van Den Hengel. Vision-and-language navigation: Interpreting visually-grounded navigation
530 instructions in real environments. In *Proceedings of Conference on Computer Vision and Pattern Recognition
(CVPR)*, 2018.

531
532

533 Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, and Cordelia Schmid. Vivit: A
534 video vision transformer. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2021.

535
536
537

538 Trevor Ashby, Braden K Webb, Gregory Knapp, Jackson Searle, and Nancy Fulda. Personalized quest and
539 dialogue generation in role-playing games: A knowledge graph-and language model-based approach. In
540 *ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 1–20, 2023.

541
542
543

544 Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and
545 Jingren Zhou. Qwen-vl: A frontier large vision-language model with versatile abilities. *arXiv preprint*
546 *arXiv:2308.12966*, 2023.

547
548
549

550 Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S
551 Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of
552 foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

553
554
555

556 Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Joseph Dabis, Chelsea Finn, Keerthana
557 Gopalakrishnan, Karol Hausman, Alex Herzog, Jasmine Hsu, et al. Rt-1: Robotics transformer for real-
558 world control at scale. *arXiv preprint arXiv:2212.06817*, 2022.

559
560
561
562

563 Crystal Chao, Maya Cakmak, and Andrea L Thomaz. Towards grounding concepts for transfer in goal learn-
564 ing from demonstration. In *2011 IEEE International Conference on Development and Learning (ICDL)*,
565 volume 2, pp. 1–6. IEEE, 2011.

566
567
568

569 Howard Chen, Alane Suhr, Dipendra Misra, Noah Snavely, and Yoav Artzi. Touchdown: Natural language
570 navigation and spatial reasoning in visual street environments. In *Proceedings of Conference on Computer
571 Vision and Pattern Recognition (CVPR)*, 2019.

572
573
574

575 Youjung Choi and Yuyan Luo. Understanding preferences in infancy. *Wiley Interdisciplinary Reviews: Cogni-
576 tive Science*, 14(4):e1643, 2023.

540 Géraldine Coppin, Sylvain Delplanque, Isabelle Cayeux, Christelle Porcherot, and David Sander. I'm no longer
 541 torn after choice: How explicit choices implicitly shape preferences of odors. *Psychological science*, 21(4):
 542 489–493, 2010.

543 Yan Ding, Xiaohan Zhang, Saeid Amiri, Nieqing Cao, Hao Yang, Andy Kaminski, Chad Esselink, and Shiqi
 544 Zhang. Integrating action knowledge and ilms for task planning and situation handling in open worlds.
 545 *Autonomous Robots*, 47(8):981–997, 2023.

546 Danny Driess, Fei Xia, Mehdi SM Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid,
 547 Jonathan Tompson, Quan Vuong, Tianhe Yu, et al. Palm-e: An embodied multimodal language model. In
 548 *Proceedings of International Conference on Machine Learning (ICML)*, 2023.

549 Seymour Epstein. Integration of the cognitive and the psychodynamic unconscious. *American psychologist*, 49
 550 (8):709, 1994.

551 Christine A Fawcett and Lori Markson. Children reason about shared preferences. *Developmental psychology*,
 552 46(2):299, 2010.

553 Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé
 554 III, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92, 2021.

555 Sarah A Gerson, Harold Bekkering, and Sabine Hunnius. Do you do as i do?: Young toddlers prefer and copy
 556 toy choices of similarly acting others. *Infancy*, 22(1):5–22, 2017.

557 Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson
 558 Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, et al. Ego4d: Around the world in 3,000 hours of egocentric
 559 video. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

560 Jiayuan Gu, Sean Kirmani, Paul Wohlhart, Yao Lu, Montserrat Gonzalez Arenas, Kanishka Rao, Wenhao Yu,
 561 Chuyuan Fu, Keerthana Gopalakrishnan, Zhuo Xu, et al. Rt-trajectory: Robotic task generalization via
 562 hindsight trajectory sketches. *arXiv preprint arXiv:2311.01977*, 2023.

563 Stefan E Huber, Kristian Kiili, Steve Nebel, Richard M Ryan, Michael Sailer, and Manuel Ninaus. Leveraging
 564 the potential of large language models in education through playful and game-based learning. *Educational
 565 Psychology Review*, 36(1):25, 2024.

566 Kaiwen Jiang, Stephanie Stacy, Chuyu Wei, Adelpha Chan, Federico Rossano, Yixin Zhu, and Tao Gao. Indi-
 567 vidual vs. joint perception: a pragmatic model of pointing as communicative smithian helping. In *Annual
 568 Meeting of the Cognitive Science Society (CogSci)*, 2021.

569 Kaiwen Jiang, Annya Dahmani, Stephanie Stacy, Boxuan Jiang, Federico Rossano, Yixin Zhu, and Tao Gao.
 570 What is the point? a theory of mind model of relevance. In *Annual Meeting of the Cognitive Science Society
 571 (CogSci)*, 2022.

572 D Kahneman. The psychology of preferences. *Scientific American*, 1982.

573 Yash Kant, Arun Ramachandran, Sriram Yenamandra, Igor Gilitschenski, Dhruv Batra, Andrew Szot, and
 574 Harsh Agrawal. Housekeep: Tidying virtual households using commonsense reasoning. In *Proceedings of
 575 European Conference on Computer Vision (ECCV)*, 2022.

576 Ivan Kapelyukh and Edward Johns. My house, my rules: Learning tidying preferences with graph neural
 577 networks. In *Conference on Robot Learning (CoRL)*, 2022.

578 Eric Kolve, Roozbeh Mottaghi, Winson Han, Eli VanderBilt, Luca Weihs, Alvaro Herrasti, Matt Deitke, Kiana
 579 Ehsani, Daniel Gordon, Yuke Zhu, et al. Ai2-thor: An interactive 3d environment for visual ai. *arXiv
 580 preprint arXiv:1712.05474*, 2017.

581 Isabel Leal, Krzysztof Choromanski, Deepali Jain, Avinava Dubey, Jake Varley, Michael Ryoo, Yao Lu, Fred-
 582 erick Liu, Vikas Sindhwani, Quan Vuong, et al. Sara-rt: Scaling up robotics transformers with self-adaptive
 583 robust attention. *arXiv preprint arXiv:2312.01990*, 2023.

584 Min Kyung Lee, Jodi Forlizzi, Sara Kiesler, Paul Rybski, John Antanitis, and Sarun Savetsila. Personalization
 585 in hri: A longitudinal field experiment. In *ACM/IEEE International Conference on Human-Robot Interaction
 586 (HRI)*, 2012.

587 Seongyun Lee, Sue Hyun Park, Seungone Kim, and Minjoon Seo. Aligning to thousands of preferences via sys-
 588 tem message generalization. *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*,
 589 37:73783–73829, 2024.

590 Stephen C Levinson. Pragmatics. *Cambridge UP*, 1983.

594 Daniel Leyzberg, Samuel Spaulding, and Brian Scassellati. Personalizing robot tutors to individuals' learning
 595 differences. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2014.

596

597 Chengshu Li, Ruohan Zhang, Josiah Wong, Cem Gokmen, Sanjana Srivastava, Roberto Martín-Martín, Chen
 598 Wang, Gabrael Levine, Michael Lingelbach, Jiankai Sun, et al. Behavior-1k: A benchmark for embodied ai
 599 with 1,000 everyday activities and realistic simulation. In *Conference on Robot Learning (CoRL)*, 2023a.

600

601 Zekun Li, Baolin Peng, Pengcheng He, Michel Galley, Jianfeng Gao, and Xifeng Yan. Guiding large language
 602 models via directional stimulus prompting. *Proceedings of Advances in Neural Information Processing
 Systems (NeurIPS)*, 36:62630–62656, 2023b.

603

604 Zoe Liberman, Katherine D Kinzler, and Amanda L Woodward. Origins of homophily: Infants expect people
 605 with shared preferences to affiliate. *Cognition*, 212:104695, 2021.

606

607 Sarah Lichtenstein and Paul Slovic. The construction of preference: An overview. *The construction of prefer-
 608 ence*, 1:1–40, 2006.

609

610 Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. Improved baselines with visual instruction tuning.
 611 In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024a.

612

613 Xiao Liu, Tianjie Zhang, Yu Gu, Iat Long Iong, Yifan Xu, Xixuan Song, Shudan Zhang, Hanyu Lai, Xinyi Liu,
 614 Hanlin Zhao, et al. Visualagentbench: Towards large multimodal models as visual foundation agents. *arXiv
 615 preprint arXiv:2408.06327*, 2024b.

616

617 Philippe Mongin. Expected utility theory. *Handbook of Economic Methodology*, pp. 342–350, 01 1997.

618

619 Yao Mu, Qinglong Zhang, Mengkang Hu, Wenhui Wang, Mingyu Ding, Jun Jin, Bin Wang, Jifeng Dai, Yu Qiao,
 620 and Ping Luo. Embodiedgpt: Vision-language pre-training via embodied chain of thought. In *Proceedings
 621 of Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

622

623 Minju Park, Sojung Kim, Seunghyun Lee, Soonwoo Kwon, and Kyuseok Kim. Empowering personalized
 624 learning through a conversation-based tutoring system with student modeling. In *ACM Conference on Hu-
 625 man Factors in Computing Systems (CHI)*, pp. 1–10, 2024.

626

627 Maithili Patel and Sonia Chernova. Proactive robot assistance via spatio-temporal object modeling. In *Confer-
 628 ence on Robot Learning (CoRL)*, 2023.

629

630 Maithili Patel, Aswin Gururaj Prakash, and Sonia Chernova. Predicting routine object usage for proactive robot
 631 assistance. In *Conference on Robot Learning (CoRL)*, 2023.

632

633 Yujia Peng, Jiaheng Han, Zhenliang Zhang, Lifeng Fan, Tengyu Liu, Siyuan Qi, Xue Feng, Yuxi Ma, Yizhou
 634 Wang, and Song-Chun Zhu. The tong test: Evaluating artificial general intelligence through dynamic em-
 635 bodied physical and social interactions. *Engineering*, 34:12–22, 2024.

636

637 Zhiliang Peng, Wenhui Wang, Li Dong, Yaru Hao, Shaohan Huang, Shuming Ma, and Furu Wei. Kosmos-2:
 638 Grounding multimodal large language models to the world. *arXiv preprint arXiv:2306.14824*, 2023.

639

640 Xavier Puig, Tianmin Shu, Shuang Li, Zilin Wang, Yuan-Hong Liao, Joshua B Tenenbaum, Sanja Fidler,
 641 and Antonio Torralba. Watch-and-help: A challenge for social perception and human-ai collaboration. In
 642 *Proceedings of International Conference on Learning Representations (ICLR)*, 2021a.

643

644 Xavier Puig, Tianmin Shu, Shuang Li, Zilin Wang, Yuan-Hong Liao, Joshua B Tenenbaum, Sanja Fidler,
 645 and Antonio Torralba. Watch-and-help: A challenge for social perception and human-ai collaboration. In
 646 *Proceedings of International Conference on Learning Representations (ICLR)*, 2021b.

647

648 Xavier Puig, Tianmin Shu, Joshua B Tenenbaum, and Antonio Torralba. Nopa: Neurally-guided online proba-
 649 bilistic assistance for building socially intelligent home assistants. In *International Conference on Robotics
 650 and Automation (ICRA)*, 2023a.

651

652 Xavier Puig, Eric Undersander, Andrew Szot, Mikael Dallaire Cote, Tsung-Yen Yang, Ruslan Partsey, Ruta
 653 Desai, Alexander Clegg, Michal Hlavac, So Yeon Min, et al. Habitat 3.0: A co-habitat for humans, avatars,
 654 and robots. In *Proceedings of International Conference on Learning Representations (ICLR)*, 2023b.

655

656 Gabriel Sarch, Zhaoyuan Fang, Adam W Harley, Paul Schyldlo, Michael J Tarr, Saurabh Gupta, and Katerina
 657 Fragiadaki. Tidee: Tidying up novel rooms using visuo-semantic commonsense priors. In *Proceedings of
 658 European Conference on Computer Vision (ECCV)*, 2022.

659

660 Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian
 661 Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, et al. Habitat: A platform for embodied ai research. In
 662 *Proceedings of International Conference on Computer Vision (ICCV)*, 2019.

648 John Scott et al. Rational choice theory. *Understanding contemporary society: Theories of the present*, 129:
 649 126–138, 2000.

650

651 Amartya Sen. Behaviour and the concept of preference. *Economica*, 40(159):241–259, 1973.

652

653 Mohit Shridhar, Jesse Thomason, Daniel Gordon, Yonatan Bisk, Winson Han, Roozbeh Mottaghi, Luke Zettle-
 654 moyer, and Dieter Fox. Alfred: A benchmark for interpreting grounded instructions for everyday tasks. In
 655 *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

656

657 Tianmin Shu and Yuandong Tian. M³ 3rl: Mind-aware multi-agent management reinforcement learning. In
 658 *Proceedings of International Conference on Learning Representations (ICLR)*, 2019.

659

660 Itamar Simonson. Will i like a “medium” pillow? another look at constructed and inherent preferences. *Journal*
 661 *of Consumer Psychology*, 18(3):155–169, 2008.

662

663 Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse
 664 Thomason, and Animesh Garg. Progprompt: Generating situated robot task plans using large language
 665 models. In *International Conference on Robotics and Automation (ICRA)*, 2023.

666

667 Paul Slovic. The construction of preference. *American Psychologist*, 50(5):364, 1995.

668

669 Chan Hee Song, Jiaman Wu, Clayton Washington, Brian M Sadler, Wei-Lun Chao, and Yu Su. Llm-planner:
 670 Few-shot grounded planning for embodied agents with large language models. In *Proceedings of Interna-*
 671 *tional Conference on Computer Vision (ICCV)*, 2023.

672

673 Haoyu Song, Wei-Nan Zhang, Jingwen Hu, and Ting Liu. Generating persona consistent dialogues by ex-
 674 ploring natural language inference. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*,
 675 2020.

676

677 Ioan A Sucan, Mark Moll, and Lydia E Kavraki. The open motion planning library. *IEEE Robotics & Automa-*
 678 *tion Magazine*, 19(4):72–82, 2012.

679

680 Akira Taniguchi, Shota Isobe, Lotfi El Hafi, Yoshinobu Hagiwara, and Tadahiro Taniguchi. Autonomous plan-
 681 ning based on spatial concepts to tidy up home environments with service robots. *Advanced Robotics*, 35
 682 (8):471–489, 2021.

683

684 Jesse Thomason, Michael Murray, Maya Cakmak, and Luke Zettlemoyer. Vision-and-dialog navigation. In
 685 *Conference on Robot Learning (CoRL)*, 2020.

686

687 Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix,
 688 Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation
 689 language models. *arXiv preprint arXiv:2302.13971*, 2023.

690

691 Yu-Min Tseng, Yu-Chao Huang, Teng-Yun Hsiao, Wei-Lin Chen, Chao-Wei Huang, Yu Meng, and Yun-
 692 Nung Chen. Two tales of persona in llms: A survey of role-playing and personalization. *arXiv preprint*
 693 *arXiv:2406.01171*, 2024.

694

695 Lei Wang and Ee-Peng Lim. Zero-shot next-item recommendation using large pretrained language models.
 696 *arXiv preprint arXiv:2304.03153*, 2023.

697

698 Jimmy Wu, Rika Antonova, Adam Kan, Marion Lepert, Andy Zeng, Shuran Song, Jeannette Bohg, Szymon
 699 Rusinkiewicz, and Thomas Funkhouser. Tidybot: Personalized robot assistance with large language models.
 700 *Autonomous Robots*, 47(8):1087–1102, 2023.

701

702 Manjie Xu, Guangyuan Jiang, Wei Liang, Chi Zhang, and Yixin Zhu. Active reasoning in an open-world
 703 environment. *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 36, 2024a.

704

705 Manjie Xu, Guangyuan Jiang, Wei Liang, Chi Zhang, and Yixin Zhu. Interactive visual reasoning under
 706 uncertainty. *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 36, 2024b.

707

708 Fan Yang, Zheng Chen, Ziyan Jiang, Eunah Cho, Xiaojiang Huang, and Yanbin Lu. Palr: Personalization aware
 709 llms for recommendation. *arXiv preprint arXiv:2305.07622*, 2023.

710

711 Rui Yang, Hanyang Chen, Junyu Zhang, Mark Zhao, Cheng Qian, Kangrui Wang, Qineng Wang, Teja Venkat
 712 Koripella, Marziyeh Movahedi, Manling Li, et al. Embodiedbench: Comprehensive benchmarking multi-
 713 modal large language models for vision-driven embodied agents. *arXiv preprint arXiv:2502.09560*, 2025.

714

715 Keunwoo Peter Yu, Zheyuan Zhang, Fengyuan Hu, and Joyce Chai. Efficient in-context learning in vision-
 716 language models for egocentric videos. *arXiv preprint arXiv:2311.17041*, 2023.

702 Luyao Yuan, Xiaofeng Gao, Zilong Zheng, Mark Edmonds, Ying Nian Wu, Federico Rossano, Hongjing Lu,
 703 Yixin Zhu, and Song-Chun Zhu. In situ bidirectional human-robot value alignment. *Science Robotics*, 7(68),
 704 2022.

705 Tao Yuan, Hangxin Liu, Lifeng Fan, Zilong Zheng, Tao Gao, Yixin Zhu, and Song-Chun Zhu. Joint inference of
 706 states, robot knowledge, and human (false-)beliefs. In *International Conference on Robotics and Automation*
 707 (*ICRA*), 2020.

708 Yuan Yuan, Huandong Wang, Jingtao Ding, Depeng Jin, and Yong Li. Learning to simulate daily activities via
 709 modeling dynamic human needs. In *Proceedings of the ACM Web Conference*, 2023.

710 Mary Zey. *Rational choice theory and organizational theory: A critique*. Sage, 1998.

711 Chi Zhang, Baoxiong Jia, Mark Edmonds, Song-Chun Zhu, and Yixin Zhu. Acre: Abstract causal reasoning
 712 beyond covariation. In *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*,
 713 2021.

714 Hongxin Zhang, Weihua Du, Jiaming Shan, Qinhong Zhou, Yilun Du, Joshua B Tenenbaum, Tianmin Shu, and
 715 Chuang Gan. zhang2023building. In *Proceedings of International Conference on Learning Representations*
 716 (*ICLR*), 2024a.

717 Kai Zhang, Yangyang Kang, Fubang Zhao, and Xiaozhong Liu. Llm-based medical assistant personalization
 718 with short-and long-term memory coordination. In *North American Chapter of the Association for Compu-
 719 tational Linguistics: Human Language Technologies (NAACL-HLT)*, pp. 2386–2398, 2024b.

720 Susan Zhang, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan,
 721 Mona Diab, Xian Li, Xi Victoria Lin, et al. Opt: Open pre-trained transformer language models. *arXiv
 722 preprint arXiv:2205.01068*, 2022.

723 Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. Dags with no tears: Continuous op-
 724 timization for structure learning. In *Proceedings of Advances in Neural Information Processing Systems*
 725 (*NeurIPS*), 2018.

726 Yixin Zhu, Chenfanfu Jiang, Yibiao Zhao, Demetri Terzopoulos, and Song-Chun Zhu. Inferring forces and
 727 learning human utilities from videos. In *Proceedings of Conference on Computer Vision and Pattern Recog-
 728 nition (CVPR)*, 2016.

729 Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan
 730 Welker, Ayzaan Wahid, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic con-
 731 trol. In *Conference on Robot Learning (CoRL)*, 2023.

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756 **A DATASET CARD**

757
758 We follow the datasheet proposed in Gebru et al. (2021) for documenting our proposed PBP:

759
760 **1. Motivation**

761 (a) **For what purpose was the dataset created?**

762 The benchmark was created to evaluate existing learning agents on their ability to understand
763 and adapt to various human preferences. Specifically, it aims to test the agents' proficiency in
764 few-shot learning from demonstrations, where they must respond to ambiguous task instruc-
765 tions and formulate adaptive task plans based on limited examples of user preferences. The
766 benchmark is designed to highlight the challenges and gaps in current AI systems' capabili-
767 ties in planning activities and abstracting human preferences, ultimately driving advancements
768 towards developing more intelligent and personalized embodied agents.

769 (b) **Who created the dataset and on behalf of which entity?**

770 N/A.

771 (c) **Who funded the creation of the dataset?**

772 N/A.

773 (d) **Any other Comments?**

774 None.

775 **2. Composition**

776 (a) **What do the instances that comprise the dataset represent?**

777 Each instance contains an egocentric video of an agent's activity, its bird's-eye-view map of the
778 position of the agent, and a frame-level textual annotation of the current action, as shown in
779 Figure 4. Additionally, we provide a rendered third-person view of the entire process.

780 (b) **How many instances are there in total?**

781 15000.

782 (c) **Does the dataset contain all possible instances or is it a sample (not necessarily random)
783 of instances from a larger set?**

784 No. The dataset contains a set of demonstrations rendered within the simulator. The users can
785 render more diverse instances if they want. We have provided the rendering instructions.

786 (d) **What data does each instance consist of?**

787 The instances that comprise the benchmark represent various types of human preferences ap-
788 plied to different tasks within a realistic embodied scene. Each instance is designed to challenge
789 the learning agents to understand and adapt to these preferences based on a few demonstra-
790 tion examples, reflecting the diverse and hierarchical nature of user preferences in real-world sce-
791 narios. See above for data details.

792 (e) **Is there a label or target associated with each instance?**

793 Yes.

794 (f) **Is any information missing from individual instances?**

795 No.

796 (g) **Are relationships between individual instances made explicit?**

797 Yes.

798 (h) **Are there recommended data splits?**

799 No.

800 (i) **Are there any errors, sources of noise, or redundancies in the dataset?**

801 No.

802 (j) **Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g.,
803 websites, tweets, other datasets)?**

804 Self-contained.

805 (k) **Does the dataset contain data that might be considered confidential (e.g., data that is pro-
806 tected by legal privilege or by doctor-patient confidentiality, data that includes the content
807 of individuals' non-public communications)?**

808 No.

809 (l) **Does the dataset contain data that, if viewed directly, might be offensive, insulting, threat-
810 ening, or might otherwise cause anxiety?**

811 No.

812 (m) **Does the dataset relate to people?**

813 No.

814 (n) **Does the dataset identify any subpopulations (e.g., by age, gender)?**

815 No.

816 (o) **Is it possible to identify individuals (i.e., one or more natural persons), either directly or
817 indirectly (i.e., in combination with other data) from the dataset?**

818 No.

810 (p) **Does the dataset contain data that might be considered sensitive in any way (e.g., data**

811 that reveals racial or ethnic origins, sexual orientations, religious beliefs, political opinions

812 or union memberships, or locations; financial or health data; biometric or genetic data;

813 forms of government identification, such as social security numbers; criminal history)?

814 No.

815 (q) **Any other comments?**

816 None.

817 **3. Collection Process**

818 (a) **How was the data associated with each instance acquired?**

819 We render PBP using NVIDIA’s Omniverse and OmniGibson simulation environment (Li et al.,

820 2023a).

821 (b) **What mechanisms or procedures were used to collect the data (e.g., hardware apparatus**

822 **or sensor, manual human curation, software program, software API)?**

823 The data for each instance in the benchmark was acquired by sampling preferences from a

824 predefined set and constructing tasks paired with a few demonstrations that shared high-level

825 preferences but differed in specific objects and scenes. Each sampled preference was randomly

826 assigned to one of the 50 scenes provided by OmniGibson, with relevant objects sampled within

827 the scene. Egocentric observation and action sequences of an embodied agent were generated as

828 the agent performed tasks guided by a rule-based planner using planning primitives like inverse

829 kinematics for grasping and the A* algorithm for movement.

830 (c) **If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deter-**

831 **ministic, probabilistic with specific sampling probabilities)?**

832 N/A.

833 (d) **Who was involved in the data collection process (e.g., students, crowdworkers, contrac-**

834 **tors) and how were they compensated (e.g., how much were crowdworkers paid)?**

835 N/A.

836 (e) **Over what timeframe was the data collected?**

837 N/A.

838 (f) **Were any ethical review processes conducted (e.g., by an institutional review board)?**

839 The dataset raises no ethical concerns.

840 (g) **Does the dataset relate to people?**

841 No.

842 (h) **Did you collect the data from the individuals in question directly, or obtain it via third**

843 **parties or other sources (e.g., websites)?**

844 N/A.

845 (i) **Were the individuals in question notified about the data collection?**

846 N/A.

847 (j) **Did the individuals in question consent to the collection and use of their data?**

848 N/A.

849 (k) **If consent was obtained, were the consenting individuals provided with a mechanism to**

850 **revoke their consent in the future or for certain uses?**

851 N/A.

852 (l) **Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a**

853 **data protection impact analysis) been conducted?**

854 Yes.

855 (m) **Any other comments?**

856 None.

857 **4. Preprocessing, Cleaning and Labeling**

858 (a) **Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucket-**

859 **ing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances,**

860 **processing of missing values)?**

861 N/A.

862 (b) **Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to**

863 **support unanticipated future uses)?**

864 N/A.

865 (c) **Is the software used to preprocess/clean/label the instances available?**

866 N/A.

867 (d) **Any other comments?**

868 None.

869 **5. Uses**

870 (a) **Has the dataset been used for any tasks already?**

871 No, the dataset is newly proposed by us.

(b) **Is there a repository that links to any or all papers or systems that use the dataset?**
No, the dataset is new.

(c) **What (other) tasks could the dataset be used for?**
This dataset could be used for research topics like embodied AI and human-computer interaction.

(d) **Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?**
N/A.

(e) **Are there tasks for which the dataset should not be used?**
N/A.

(f) **Any other comments?**
None.

Distribution

(a) **Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?**
No before it is made public.

(b) **How will the dataset be distributed (e.g., tarball on website, API, GitHub)?**
On our project website upon acceptance.

(c) **When will the dataset be distributed?**
Upon acceptance.

(d) **Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?**
Under CC BY-NC¹ license.

(e) **Have any third parties imposed IP-based or other restrictions on the data associated with the instances?**
No.

(f) **Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?**
No.

(g) **Any other comments?**
None.

7. Maintenance

(a) **Who is supporting/hosting/maintaining the dataset?**
The authors.

(b) **How can the owner/curator/manager of the dataset be contacted (e.g., email address)?**
N/A.

(c) **Is there an erratum?**
Future erratum will be released through the website.

(d) **Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances')?**
Yes.

(e) **If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were individuals in question told that their data would be retained for a fixed period of time and then deleted)?**
N/A. The dataset does not relate to people.

(f) **Will older versions of the dataset continue to be supported/hosted/maintained?**
Yes.

(g) **If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?**
Yes. We will release the source code as well as a licence on our project website after acceptance.

(h) **Any other comments?**
None.

B DATASET STATISTICS

The length of the simulations in the dataset ranges from 1 to 5 minutes, depending on the tasks recorded. And the videos are recorded at 30 fps.

¹<https://creativecommons.org/licenses/by-nc/4.0/>

918 B.1 PREFERENCES
919920 See Table A1 for the preference statistics in PbP.
921922 Table A1: Dataset Statistics in PbP.
923

	Action_Level	Option_Level	Sequence_Level
Preference Num	75	135	80
Video Num	5000	5000	5000
Sub-task Num	1	2-3	2-3

928
929 B.2 ACTIONS
930931 See Table A2 for the action statistics in PbP. We implement 17 action primitives in PbP to assist with model
932 planning and dataset rendering. These action primitives have parameters that simplify tasks and are considered
933 the lowest-level actions. Each sub-task contains 8 to 20 such lowest-level actions. Generally, most of these
934 actions consist of two parts: the robot movement part and the arm (gripper) execution part. For robot move-
935 ment, we use the A* algorithm to find paths and avoid collisions. We build a connection map during scene
936 initialization for navigation, taking the robot's width into consideration. For the arm (gripper) execution, we
937 primarily use the IK algorithm to compute arm movements. However, since IK cannot handle complex tasks,
938 such as picking objects from the fridge, we also leverage the Open Motion Planning Library (OMPL) planner
(Sucan et al., 2012) with forward planning to assist in planning the arm positions.
939940 B.3 MORE DATASET DETAILS AND DISCUSSION
941942 **Dataset production** The process of producing data is mainly explained in Section 4.2. In summary, we
943 follow the order of “sample preference - sample scene - sample objects to be manipulated - generate actions
944 guided by a rule-based planner”.945 **Length and FPS of the simulations** The length of the simulations ranges from 1 to 5 minutes, depending
946 on the tasks recorded. The videos are recorded at 30 fps.
947948 **Actions contained in each simulation** The number of actions in simulations varies among different pref-
949 erence levels. There is 1 subtask for action-level, 2-3 subtasks for option-level, and 2-3 subtasks for sequence-
950 level preferences. Each subtask contains 8-20 actions.
951952 **Scenes and rooms** Each scene contains various types of rooms. The main differences between scenes are
953 the type, number, and layout of both rooms and furniture. Additionally, each room may contain different objects
954 and have unique layouts. Details of the scenes and rooms can be found in Omnidigibson’s official documentation
955 (<https://behavior.stanford.edu/omnidigibson/>), as we directly adopt these scenes from the open-sourced project.
956957 **290 preference types** Considering that preferences in household activities are not only multi-dimensional
958 but also hierarchical, we first define a hierarchy of preferences from the perspective of how things happen in
959 a life scenario, that is, from each specific action to a sub-task consisting of several actions, and then to the
960 sequence combining these sub-tasks. The next step is to expand each level with typical tasks and actions. The
961 detailed definition of the 290 preferences can be found in Section 4.1.
962963 **The egocentric view** Collecting both egocentric observations and third-person views is feasible in PbP or
964 similar environments built on simulators like iGibson. However, in real-world scenarios, it is generally easier to
965 gather egocentric observations of human daily activities, as these can be efficiently captured through wearable
966 devices. Additionally, there are numerous egocentric-view datasets available, such as Ego4D(Grauman et al.,
967 2022), which further facilitate this approach. While third-person views can provide a different perspective,
968 they often encounter issues such as occlusion. Although research based on third-person views is essential
969 for applications involving real robots, focusing on egocentric views in the current work allows for a more
970 straightforward exploration of preference learning and planning. Nevertheless, third-person view data can be
971 obtained by integrating additional cameras, as outlined in our provided code.972 **Action ground truth** In experiments involving vision input, we do not explicitly provide the action se-
973 quence of the user. In the symbolic-based experiment, we provide the action sequence to reduce the perception
974 cost to concentrate more effectively on the inference and planning aspects of the study.
975

972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989

Table A2: Action Primitives in PBP.

Action List	Explanation
Move_to_[]	Move to a specified location, or a specified room, or a specified object
Rotate_to_[]	Rotate to a specified orientation or a specified object
Pick_[]	Pick up an object using the gripper, e.g., “Pick_apple”
Place_[]	Place an object at a location, e.g., “Place_apple_on_table”
Fill_[]_with_[]	Fill a container with a substance, e.g., “Fill_glass_with_water”
Pour_[]	Pour a substance from a container, e.g., “Pour_milk”
Open_[]	Open an object, e.g., “Open_door”
Close_[]	Close an object, e.g., “Close_fridge”
Cut_[]	Cut an object, e.g., “Cut_carrot”
Cook_[]	Cook an item, e.g., “Cook_pasta”
Wash_[]	Wash an object, e.g., “Wash_dishes”
Clean_[]	Clean a surface or object, e.g., “Clean_counter”
Cover_[]	Cover an object, e.g., “Cover_bowl”
Uncover_[]	Uncover an object, e.g., “Uncover_bowl”
Toggle_on_[]	Turn on a device, e.g., “Toggle_on_light”
Toggle_off_[]	Turn off a device, e.g., “Toggle_off_stove”
Wait_[]	Wait some time

C EXPERIMENT DETAILS

C.1 CASE STUDY

We also provide a case with preference *Put fruit on the bed* in the following table Appendix C.1. We present a simplified version of the demonstrations, where all video outputs have been translated into symbol-based action sequences for ease of understanding. Video-based models such as LLaVA-Next and GPT-4V struggle with comprehending preferences and tend to replicate certain action patterns from the video demonstration, such as “move to” and “pick up”. Llama3 demonstrates a partial understanding and execution of the preference. It correctly moves to each fruit (grape, banana), picks them up, and places them on the bed. However, it also interacts with the pencil and places it on the bed, which is not required by the preference. Ideally, the pencil should be placed on the table, similar to the pen. On the other hand, GPT-4.1(Symbol) accurately interacts with the grape and banana by moving to each fruit, picking it up, and placing it on the bed. This demonstrates a better understanding and execution of the preference compared to the other models.

D BASELINE DETAILS

D.1 ViViT

As a baseline, we employ the pure-Transformer-based Video Vision Transformer (ViViT) (Arnab et al., 2021), an end-to-end trainable model with proven capabilities in extracting spatial and temporal information from video inputs. Since it lacks a LLM component, ViViT likely serves as a lower bound for commonsense understanding in PBP tasks.

Inspired by Vision Transformer, ViViT extracts spatio-temporal tokens from the input video and outputs video classification labels for classification. We adopt the ViViT implementation from the official GitHub repo <https://github.com/google-research/scenic>.

Specifically, we utilize a ViViT with an image size of 224 and a patch size of 16. We extract 2 frames per second from the input video and pad them with the last frame. The Transformer architecture with 3 attention heads operates on features of hidden size of 192 and depth of 4. Each attention head operates on a dimension of 64. We train our model for 30 epochs with a learning rate 3e-5. For the few-shot setting, we concatenate the demo videos temporally.

D.2 LLaVA

Building on more sophisticated architectures, LLaVA (Liu et al., 2024a) represents an end-to-end trainable large multimodal model that integrates vision and text for comprehensive visual-language understanding. We specifically evaluate LLaVA-NeXT, which has been finetuned to excel at zero-shot video understanding tasks.

Following the official implementation of LLaVA from <https://github.com/LLaVA-VL/LLaVA-NeXT>, we test the LLaVA-NeXT-Video-7B-DPO model which is designed for video under-

Case	ID	Scene	Actions
Context	Demo 1	Beechwood	move to orange, pick up orange, move to bed with orange, place orange on bed, move to apple, pick up apple, move to bed with apple, place apple on bed, move to dvd, pick up dvd, move to table with dvd, place dvd on table.
	Demo 2	Rs	move to banana, pick up banana, move to bed with banana, place banana on bed, move to club sandwich, pick up club sandwich, move to countertop with club sandwich, place club sandwich on countertop, move to orange, pick up orange, move to countertop with orange, move to bed with orange, place orange on bed.
	Demo 3	Merom	move to pen, pick up pen, move to table with pen, place pen, move to strawberry, pick up strawberry, move with to bed with strawberry, place strawberry on bed, move to textbook, pick up textbook, move with textbook, move to bookshelf with textbook, place textbook in bookshelf.
Test Scene		Beechwood	grape, pencil, banana
Answer	LLaVA-Next	Beechwood	move to grape, move to bed with grape, move to pencil, pick up pencil, move to banana, pick up banana
	GPT-4V	Beechwood	move to grape, pick up grape, move to bed with grape, place grape on bed
	Llama3	Beechwood	move to grape, pick up grape, move to bed with grape, place grape on bed, move to pencil, pick up pencil, move to bed with pencil, place pencil on bed, move to banana, pick up banana, move to bed with banana, place banana on bed
	GPT-4.1(Symbol)	Beechwood	move to grape, pick up grape, move to bed with grape, place grape on bed, move to banana, pick up banana, move to bed with banana, place banana on bed

Table A3: Case Study with preference *Put fruit on the bed*.

standing. Specifically, we run the model following the default inference settings, with vicuna_v1 as the prompt mode, a sample frame number of 32, and a spatial pooling stride of 2. The textual prompts are as follows²:

“Stage One / Preference Prediction”

You are a robot assistant that can help summarize the host's preference.

All possible preferences are: {ALL POSSIBLE PREFERENCES}

Now there are some previous video demos:

[VIDEO_DEMO_1] The preference is [PREFERENCE_1]

[VIDEO_DEMO_2] The preference is [PREFERENCE_2]

[VIDEO_DEMO_3] The preference is [PREFERENCE_3]

Now, please summarize the preference from the last video: [TEST_CASE]

Quesiton: What's the user's preference? Choose from the preference listed
→ before:

“Stage Two / Planning”

You are a robot assistant. Please view the demos and help generate action
→ sequence.

All possible preferences are: {ALL POSSIBLE ACTIONS}

Now there are some previous video demos:

[VIDEO_DEMO_1]

[VIDEO_DEMO_2]

²For the textual prompts, we aim to maintain consistency across all LLMs, although some baselines may have additional requirements for the input format. The prompt design is mainly motivated by OpenAI Cookbook [git@github.com:openai/openai-cookbook.git](https://github.com/openai/openai-cookbook.git). We omitted the prompt tuning process, as we found that minor changes in the prompt were unlikely to significantly impact the results. Conversely, selecting the proper demonstrations in the few-shot examples has a much greater influence on the results.

1080 [VIDEO_DEMO_3]
 1081 Now you are in the scene with [SCENE DESCRIPTIONS]. Your action sequence
 1082 → is:
 1083

1084 **D.3 EILEV**

1085 For specialized egocentric video processing, we incorporate EILEV (Yu et al., 2023), which achieves in-context
 1086 learning through architectural modifications to a pretrained VLM. Our implementation uses OPT-2.7B (Zhang
 1087 et al., 2022) as the language backbone. The model’s pretraining on Ego4D (Grauman et al., 2022) aligns well
 1088 with PBP’s egocentric perspective.

1089 Following the official implementation from <https://github.com/yukw777/EILEV.git>, we test the
 1090 EILEV model in PBP. There are two reasons we chose EILEV among other VLMs as one of our baselines: 1)
 1091 EILEV elicits in-context learning through a series of architectural modifications and a unique training process,
 1092 2) EILEV is trained using ego-centric data, which is compatible with PBP’s input. The textual prompts are as
 1093 follows. Since EILEV requires the input of the videos and texts to follow a certain pattern for better in-context
 1094 learning, there are some small modifications to the prompt:

1095 “Stage One / Preference Prediction”
 1096 You are a robot assistant that can help summarize the host’s preference.
 1097 All possible preferences are: {ALL POSSIBLE PREFERENCES}
 1098 Quesiton: What’s the user’s preference? Choose from the preference listed
 → before:
 1099 Now there are some previous video demos:
 1100 [VIDEO_DEMO_1] The preference is [PREFERENCE_1]
 1101 [VIDEO_DEMO_2] The preference is [PREFERENCE_2]
 1102 [VIDEO_DEMO_3] The preference is [PREFERENCE_3]
 1103 [TEST_CASE]

1104 “Stage Two / Planning”
 1105 You are a robot assistant. Please view the demos and help generate action
 → sequence.
 1106 All possible preferences are: {ALL POSSIBLE ACTIONS}
 1107 Now there are some previous video demos:
 1108 [VIDEO_DEMO_1]
 1109 [VIDEO_DEMO_2]
 1110 [VIDEO_DEMO_3]
 1111 Now you are in the scene with [SCENE DESCRIPTIONS]. Your action sequence
 → is:
 1112

1113

1114 **D.4 GPT-4V**

1115 To benchmark against state-of-the-art visual-language models, we evaluate GPT-4V using the Azure OpenAI
 1116 API (version “gpt-4-turbo-2024-04-09”). Due to image token limitations, we implement video input subsam-
 1117 pling while maintaining temporal coherence. The API has a limit of 10 images per request. Consequently, for
 1118 the zero-shot setting, we resample each input video to 8 frames of size 224. For the few-shot setting, where we
 1119 need to input 3 extra video demonstrations, we concatenate 4 images into a frame, thereby obtaining 4 videos
 1120 in 8 frames, maintaining the same frame number as the previous setting. We test the model with a temperature
 1121 of 0.05. The textual prompts are as follows:

1122 “Stage One / Preference Prediction”
 1123 You are a robot assistant that can help summarize the host’s preference.
 1124 All possible preferences are: {ALL POSSIBLE PREFERENCES}
 1125 Now there are some previous video demos:
 1126 [VIDEO_DEMO_1] The preference is [PREFERENCE_1]
 1127 [VIDEO_DEMO_2] The preference is [PREFERENCE_2]
 1128 [VIDEO_DEMO_3] The preference is [PREFERENCE_3]
 1129 Now, please summarize the preference from the last video: [TEST_CASE]
 1130 Quesiton: What’s the user’s preference? Choose from the preference listed
 → before:
 1131 “Stage Two / Planning”
 1132 You are a robot assistant. Please view the demos and help generate action
 → sequence.
 1133 All possible preferences are: {ALL POSSIBLE ACTIONS}

1134 Now there are some previous video demos:
 1135 [VIDEO_DEMO_1]
 1136 [VIDEO_DEMO_2]
 1137 [VIDEO_DEMO_3]
 1138 Now you are in the scene with [SCENE DESCRIPTIONS]. Your action sequence
 → is:
 1139

1140 Beyond multimodal approaches, we also evaluate single-modal models that process only action sequences:
 1141

1142 D.5 DAG-OPT

1143 We approach symbolic reasoning by framing the problem as a DAG-Optimization task that uncovers dependency relations between actions and preferences (Zheng et al., 2018). Our implementation uses a score-based
 1144 NOTEARS model to learn a generalized Structural Equation Model (SEM), following previous few-shot reasoning frameworks (Zhang et al., 2021; Xu et al., 2024b) based on causal dependency structures.
 1145

1146 We implement the DAG-Opt baseline following <https://github.com/xunzheng/notears.git>. Specifically, we implement a nonlinear NOTEARS using MLP in evaluation.
 1147

1148 To assess pure language understanding, we evaluate advanced LLMs including Llama3 (Touvron et al., 2023) and GPT-4.1 (Achiam et al., 2023) using only action sequence inputs. This approach treats actions as high-level abstractions of egocentric videos, reducing visual complexity while maintaining task semantics. We benchmark Llama3-8B as our baseline against GPT-4.1 as the current state-of-the-art, employing prompt designs informed by the OpenAI Cookbook for optimal few-shot performance.
 1149

1150 D.6 LLAMA3-8B

1151 We test the Llama3 series model with the official scripts from <https://github.com/meta-llama/llama3>. Specifically, we test the 8B instruction-tuned variant “Meta-Llama-3-8B-Instruct” on PBP. We test the model with a temperature of 0.05. The textual prompts are as follows:
 1152

1153 “Stage One / Preference Prediction”
 1154 You are a robot assistant that can help summarize the host's preference.
 1155 Please read the following text file and summarize the user's preference.
 1156 All possible preferences are: {ALL POSSIBLE PREFERENCES}
 1157 [TEXT_ANNOTATION_1] The preference is [PREFERENCE_1]
 1158 [TEXT_ANNOTATION_2] The preference is [PREFERENCE_2]
 1159 [TEXT_ANNOTATION_3] The preference is [PREFERENCE_3]
 1160 Now, please summarize the preference from the last tet file: [TEST_CASE]
 1161 Qusiton: What's the user's preference? Choose from the preference listed
 → before:
 1162

1163 “Stage Two / Planning”
 1164 You are a robot assistant. Please read the following text files and help
 → generate action sequence.
 1165 All possible preferences are: {ALL POSSIBLE ACTIONS}
 1166 Now there are some previous video demos:
 1167 [TEXT_ANNOTATION_1] (action sequence)
 1168 [TEXT_ANNOTATION_2] (action sequence)
 1169 [TEXT_ANNOTATION_3] (action sequence)
 1170 Now you are in the scene with [SCENE DESCRIPTIONS]. Your action sequence
 → is:
 1171

1172 D.7 DEEPSEEK-R1

1173 We use a self-hosted DeepSeek-R1-671B model with a temperature setting of 0.05. For our textual prompts, we deliberately disable the deep thinking capability by inserting an empty thinking section (“<think>\n\n</think>”) into the assistant content, as our experiments have shown that the deep thinking feature provides no performance improvement for tasks in PBP:
 1174

1175 “Stage One / Preference Prediction”
 1176 You are a robot assistant that can help summarize the host's preference.
 1177 Please read the following text file and summarize the user's preference.
 1178 All possible preferences are: {ALL POSSIBLE PREFERENCES}
 1179 [TEXT_ANNOTATION_1] The preference is [PREFERENCE_1]
 1180 [TEXT_ANNOTATION_2] The preference is [PREFERENCE_2]

1188 [TEXT_ANNOTATION_3] The preference is [PREFERENCE_3]
 1189 Now, please summarize the preference from the last tet file: [TEST_CASE]
 1190 Quesiton: What's the user's preference? Choose from the preference listed
 1191 → before:
 1192 "Stage Two / Planning"
 1193 You are a robot assistant. Please read the following text files and help
 1194 → generate action sequence.
 1195 All possible preferences are: {ALL POSSIBLE ACTIONS}
 1196 Now there are some previous video demos:
 1197 [TEXT_ANNOTATION_1] (action sequence)
 1198 [TEXT_ANNOTATION_2] (action sequence)
 1199 [TEXT_ANNOTATION_3] (action sequence)
 1200 Now you are in the scene with [SCENE DESCRIPTIONS]. Your action sequence
 1201 → is:
 1202 <think>\n\n</think>

D.8 GPT-4.1

We use “gpt-4.1-2025-04-14” with a temperature of 0.05. The textual prompts are as follows:

1204 "Stage One / Preference Prediction"
 1205 You are a robot assistant that can help summarize the host's preference.
 1206 Please read the following text file and summarize the user's preference.
 1207 All possible preferences are: {ALL POSSIBLE PREFERENCES}
 1208 [TEXT_ANNOTATION_1] The preference is [PREFERENCE_1]
 1209 [TEXT_ANNOTATION_2] The preference is [PREFERENCE_2]
 1210 [TEXT_ANNOTATION_3] The preference is [PREFERENCE_3]
 1211 Now, please summarize the preference from the last tet file: [TEST_CASE]
 1212 Quesiton: What's the user's preference? Choose from the preference listed
 1213 → before:
 1214 "Stage Two / Planning"
 1215 You are a robot assistant. Please read the following text files and help
 1216 → generate action sequence.
 1217 All possible preferences are: {ALL POSSIBLE ACTIONS}
 1218 Now there are some previous video demos:
 1219 [TEXT_ANNOTATION_1] (action sequence)
 1220 [TEXT_ANNOTATION_2] (action sequence)
 1221 [TEXT_ANNOTATION_3] (action sequence)
 1222 Now you are in the scene with [SCENE DESCRIPTIONS]. Your action sequence
 1223 → is:

E FUTURE DISCUSSION

E.1 LIMITATIONS AND FUTURE WORK

Our work's primary limitation stems from its reliance on synthetic data. Despite Omniverse's high-quality scene rendering, the simulator cannot fully replicate real-world complexity and variability. Furthermore, human-defined preference labels may not completely capture preference subtleties and diversity. We are addressing these limitations by collecting real-world preference demonstrations using head-worn devices.

1231
 1232
 1233
 1234
 1235
 1236
 1237
 1238
 1239
 1240
 1241