Communication-Efficient Desire Alignment for Proactive Embodied Human–Agent Interaction

Yuanfei Wang[1,2], Xinju Huang[3,5,6], Fangwei Zhong[3,5,6†], Yaodong Yang[2,4], Yizhou Wang[1,4], Yuanpei Chen[2,4†], Hao Dong[1†]
[1]School of Computer Science, Peking University [2]PKU-PsiBot Joint Lab
[3]School of Artificial Intelligence, Beijing Normal University
[4]Institute for Artificial Intelligence, Peking University
[5]Beijing Key Laboratory of Artificial Intelligence for Education
[6]Engineering Research Center of Intelligent Technology and Educational Application

Abstract

Effective real-world human–agent interactions, such as household robotic services, are often long-term and repeated. Beyond executing tasks, agents are expected to quickly become familiar with individual users. In everyday use, people do not want to repeatedly specify precise instructions. Instead, they prefer agents that adapt to their habits and preferences over interaction while minimizing communication effort. This poses a key challenge: enabling agents to rapidly align with user needs and provide proactive assistance within limited communication. To study this problem in a realistic embodied setting, we first introduce HA-Desire, a home assistance simulation environment. HA-Desire features an LLM-driven proxy user with value-driven preferences and natural language behavior, enabling systematic evaluation of how agents adapt to users across interactions and satisfy their desires. We further propose FAMER, a framework that integrates goal-relevant memory, desire-centered mental reasoning, and efficient communication to infer user preferences from interaction while reducing unnecessary dialogue. Experiments across embodied household tasks and different LLMs show that FAMER improves both task success and interaction efficiency compared to existing baselines, highlighting the importance of communication-efficient desire alignment for proactive embodied agents that support users without requiring frequent instructions.

1 Introduction

Human–agent interaction requires agents to collaborate with people in a natural, efficient, and user-centered manner. In many real-world applications, interaction is inherently long-term and repeated. Users expect agents to gradually become familiar with their habits, preferences, and needs, rather than requiring precise instructions for every task.

† Corresponding Authors

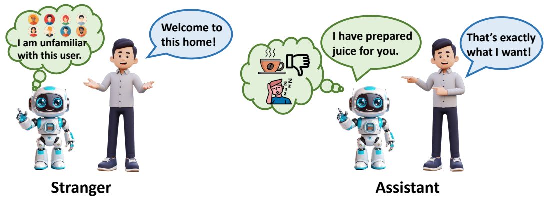

Figure 1: An illustration of proactive communication-efficient human–agent interaction in a household setting. A home assistant robot initially interacts with an unfamiliar user. Over interactions, the agent gradually models the user's habits and preferences, such as avoiding caffeine and preferring refreshing drinks in the morning due to poor sleep. Without requiring precise instructions, the agent proactively serves juice, demonstrating effective user adaptation and high-quality service.

Frequent clarification or excessive communication can interrupt interaction and increase user burden. This problem is particularly challenging for embodied agents operating in physical environments, which must not only interpret user desires but also perceive complex scenes, reason about objects and actions, and execute multi-step tasks. At the same time, users interacting with embodied agents often provide brief or underspecified instructions, expecting the agent to adapt through interaction and offer proactive assistance. As a result, enabling fast and communication-efficient alignment with individual users is a central challenge for embodied human–agent interaction.

Home assistant robots exemplify this challenge. Even when trained on broad human-centric datasets, an agent inevitably encounters unfamiliar users whose preferences and values are unknown at deployment. To be helpful, the agent must adapt through interaction, minimizing repetitive communication while improving its understanding of the user. Ideally, the agent should behave like a considerate assistant that learns from experience and anticipates user needs. As illustrated in Figure 1, a robot entering a new home gradually learns that the

Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 14094–14108
July 2-7, 2026 ©2026 Association for Computational Linguistics

user avoids caffeine and prefers refreshing breakfast options due to poor sleep. Rather than repeatedly asking for instructions, the robot proactively serves juice, demonstrating successful alignment with the user's preferences.

Prior work has explored collaboration with unfamiliar partners through paradigms such as ad-hoc teamwork (Rahman et al., 2021; Ravula, 2019; Stone et al., 2010) and zero-shot coordination (Cui et al., 2021; Hu et al., 2021, 2020). While these approaches address coordination under limited prior knowledge, they are typically studied in simplified domains such as board games (Bard et al., 2020) or 2D grid-based environments (Albrecht and Ramamoorthy, 2015; Carroll et al., 2019). These settings lack realistic embodied actions, natural language interaction, and human-like, value-driven preferences, limiting their ability to capture the challenges of long-term adaptation in embodied human–agent interaction.

To address this gap, we introduce HA-Desire (Home Assistance with Diverse Desire), an embodied simulation environment built on VirtualHome (Puig et al., 2018). HA-Desire provides rich 3D household scenes, everyday tasks such as preparing snacks or setting a table, and an LLM-driven proxy user that exhibits value-driven preferences and communicates in natural language. Rather than repeatedly specifying goals, the proxy user provides preference-oriented instructions, reflecting realistic interaction where users aim to minimize communication effort. This design enables systematic evaluation of how efficiently an embodied agent can adapt to a user across interactions.

Within HA-Desire, we further propose FAMER (**F**ast **A**daptation via **ME**ntal **R**easoning), an LLM-driven framework for communication-efficient desire alignment. FAMER integrates three key components: (1) desire-centered mental reasoning to infer user preferences from interaction signals, (2) reflection-based efficient communication to reduce redundant dialogue, and (3) goal-relevant information extraction with cross-episode persistent memory to enable reuse of previously acquired knowledge. Together, these components allow an agent to rapidly model user preferences and provide proactive assistance in complex embodied tasks.

We evaluate FAMER on representative household tasks in HA-Desire under varying goal complexities. Experiments with both LLM-driven proxy users and real human users show that FAMER significantly improves task success and communication efficiency compared to baselines. Ablation studies further confirm the contribution of each component, highlighting the importance of communication-efficient desire alignment for proactive embodied human–agent interaction.

In summary, our contributions include: (1) We formulate the problem of communication-efficient adaptation to unfamiliar, value-driven users in embodied human–agent interaction, and introduce HA-Desire, a 3D home assistance simulation environment with naturalistic user behavior for systematic evaluation. (2) We propose FAMER, a new framework that integrates desire-centered mental reasoning, reflection-based efficient communication, and goal-relevant information memory to enable fast alignment with user preferences in complex embodied tasks. (3) We validate the effectiveness of both HA-Desire and FAMER through extensive quantitative and qualitative experiments, demonstrating improved task success and interaction efficiency compared to existing baselines.

2 Related Work

Value Alignment has been extensively studied in both language models and agent design. In the context of LLMs, alignment techniques such as RLHF (Ouyang et al., 2022; Dai et al., 2023; Ji et al., 2023) aim to align models with human preferences, but these efforts primarily focus on static, text-based tasks and do not address the challenges of dynamic, embodied interactions. In human-AI collaboration, value alignment involves inferring user preferences through feedback (Yuan et al., 2022; Hiatt et al., 2017; Fisac et al., 2020). More closely related to our setting are mental reasoning agents inspired by Theory of Mind (Rabinowitz et al., 2018; Wang et al., 2022), which model other agents' beliefs and desires to support assistance. D2A (Wang et al., 2025) simulates human desires using LLMs, but is limited to text-based tasks. CHAIC (Du et al., 2024) introduces an embodied social intelligence challenge that focuses on reasoning under physical constraints, but does not address the diversity of human values and goals. In contrast, our work introduces an embodied simulation platform with naturalistic, value-driven goal generation and communication.

Adaptive Agents. Adaptation in multi-agent settings has been studied under the paradigms of zero-shot coordination (ZSC) (Hu et al., 2020, 2021; Cui et al., 2021; Lupu et al., 2021; Strouse et al., 2021)

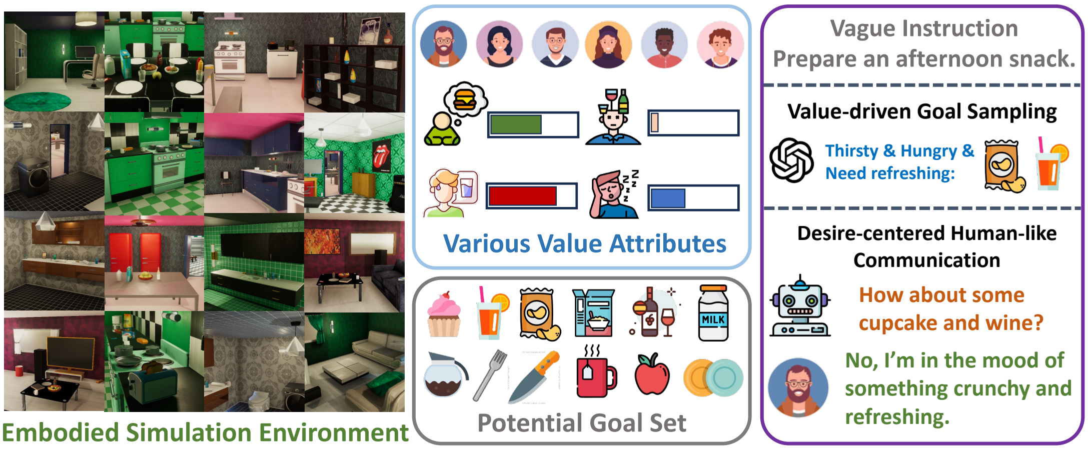

Figure 2: Overview of the HA-Desire environment. The simulation contains diverse objects and scenes. A proxy human user is associated with latent value attributes that guide goal selection from a potential goal set via an LLM. Instead of explicitly specifying goals, the user communicates through preference-oriented, natural language hints. This setup enables systematic evaluation of how embodied agents infer user preferences through interaction and adapt their behavior while minimizing communication.

and ad-hoc teamwork (AHT) (Stone et al., 2010; Rahman et al., 2021; Chen et al., 2020; Mirsky et al., 2020; Ma et al., 2024), where agents must coordinate with unseen partners without prior agreement. While these approaches are effective in structured domains such as Hanabi (Bard et al., 2020) and Overcooked (Carroll et al., 2019), they rely on symbolic observations, making them less suitable for complex, embodied human-agent collaboration. More recently, LLM-based agent frameworks (Li et al., 2023; Yao et al., 2023; Zhang et al., 2024b,a; Liu et al., 2025) have demonstrated impressive capabilities in reasoning and planning within interactive environments. However, most assume known goals or static user preferences and lack mechanisms for inferring latent values through interaction. Our work instead addresses the challenge of fast adaptation to unknown, value-driven user goals via desire inference, memory utilization, and efficient, human-like dialogue in rich embodied tasks.

LLM-Agent Simulation. Recent work uses LLMs to build interactive human-like agents with memory and communication. Generative Agents (Park et al., 2023) introduced memory, reflection, and planning for believable social simulacra. Social reasoning has been benchmarked in Sotopia (Zhou et al., 2023) and extended to repeated interactions in Lifelong-Sotopia (Goel and Zhu, 2025). Human simulation has also scaled from individualized behavior modeling in Generative Agent Simulations of 1,000 People (Park et al., 2024) to societal-level simulation in AgentSociety (Piao et al., 2025). Different from these works, our setting focuses on value-driven embodied human user simulation, where a proxy user reveals latent values through repeated embodied interaction and communication, enabling the study of communication-efficient desire alignment in embodied tasks.

3 Embodied Home Assistance Simulation Environment

In real-world interaction, human often provide brief or underspecified instructions, either because their goals are still being formed or because they wish to minimize communication effort (Bettman et al., 1998). Repeated clarification can interrupt interaction and increase user burden (Clark and Brennan, 1991). As a result, an effective assistant agent should be able to gradually model user preferences through interaction and adapt its behavior accordingly, reducing unnecessary communication while providing proactive assistance.

To study communication-efficient user adaptation in embodied human–agent interaction, we introduce **HA-Desire** (Home Assistance with Diverse Desire), a simulation environment designed around realistic household assistance scenarios. As shown in Figure 2, HA-Desire builds upon VirtualHome (Puig et al., 2018) and extends it with value-driven proxy human users. The environment includes six distinct home layouts with common rooms such as kitchens, living rooms, bathrooms, and bedrooms, containing on average 80 objects per room from over 110 object classes. This diversity enables visually grounded, multi-step household tasks and supports systematic evaluation of how embodied agents adapt to users across interactions.

We describe the problem formulation and proxy human user model in more detail below.

3.1 Problem Formulation

In HA-Desire, we study communication-efficient human-agent interaction. The interaction starts from a vague task description, reflecting realistic scenarios where users provide coarse instructions and expect agents to align over time. Formally, there exists a potential goal set G_p, from which the user's actual goals are selected.

Given a task description T, the proxy human user H samples a set of latent value attributes V from a task-specific value space, and then selects a set of desire-driven goals $G = G(V, G_p, T) \subset G_p$. The goal set G represents the user's preferences in the current interaction but is not directly observable by the ego agent.

The ego agent E and the human user can communicate by exchanging messages M_E and M_H. The agent executes actions according to a policy $\pi(A \mid O, M_H, C, T)$, where O denotes the current observation and C is a cross-episode memory context that stores information from past interactions. The agent receives positive reward for completing goals aligned with the user's preferences and penalties for irrelevant actions, reflecting misalignment.

HA-Desire supports repeated interaction with the same user over K episodes, where the user's value attributes V remain fixed. This setting encourages the agent to gradually build an internal user model and reuse previously acquired knowledge. The agent's objective is to maximize cumulative score and interaction efficiency over these episodes, i.e., $\max_\pi \mathbb{E}\left[\sum_{k=1}^{K} \sum_{t=1}^{t_k} r_t^{(k)}\right]$, where $r_t^{(k)}$ denotes the score at time step t in episode k. This formulation explicitly evaluates an agent's ability to achieve fast desire alignment and provide proactive assistance under limited communication.

3.2 Value-driven Human User

As illustrated in Figure 2, the proxy human user in HA-Desire is associated with value attributes V sampled from a task-specific value space. These attributes represent stable user preferences. For example, in the *Prepare Snack* task, the value space includes dimensions such as *Hungry*, *Thirsty*, *SweetTooth*, *Fruitarian*, and *Alcoholic*, each taking one of three levels: *Not*, *Somewhat*, or *Very*.

Conditioned on the sampled values V and the vague task description T, an LLM is used to generate a corresponding set of desire-related goals $G \subset G_p$. Since multiple goals may be compatible with the same preference configuration, this goal sampling process is intentionally non-deterministic, capturing the variability and ambiguity of human decision-making in everyday interaction.

To simulate natural human-agent interaction, the proxy user communicates through preference-oriented natural language rather than explicitly specifying goals. This design choice is not intended to model deliberate goal hiding. Instead, it reflects common household interaction where "lazy" users provide high-level or even no guidance and expect agents to infer details and adapt their behavior over time. By constraining explicit goal revelation, HA-Desire enables systematic evaluation of how effectively an embodied agent can infer user preferences from interaction and deliver proactive assistance while minimizing communication effort. Details of the LLM prompts and response constraints used to generate user behavior are provided in Appendix A.

4 Efficient Human-Agent Interaction

To enable fast adaptation to users and efficient human-agent interaction, we propose a novel framework, **FAMER** (Fast Adaptation via MEntal Reasoning), which leverages the reasoning capabilities and commonsense embedded in LLMs. In this work, we utilize GPT-4o (Hurst et al., 2024).

As illustrated in Figure 3, FAMER integrates three core components: Key Information Extraction, Desire-Centered Mental Reasoning, and Efficient Communication. These modules work in concert to help the ego agent infer user intent, plan accurately, and minimize redundant interactions and communication. We detail each component in the following subsections.

4.1 Information Extraction

In HA-Desire, the ego agent receives first-person RGB-D images as observations. To extract structured information from these inputs, we employ a perception module based on Mask R-CNN (He et al., 2017), trained on collected scene images following (Zhang et al., 2024b). The module first predicts instance segmentation masks from the RGB image and then constructs 3D point clouds using

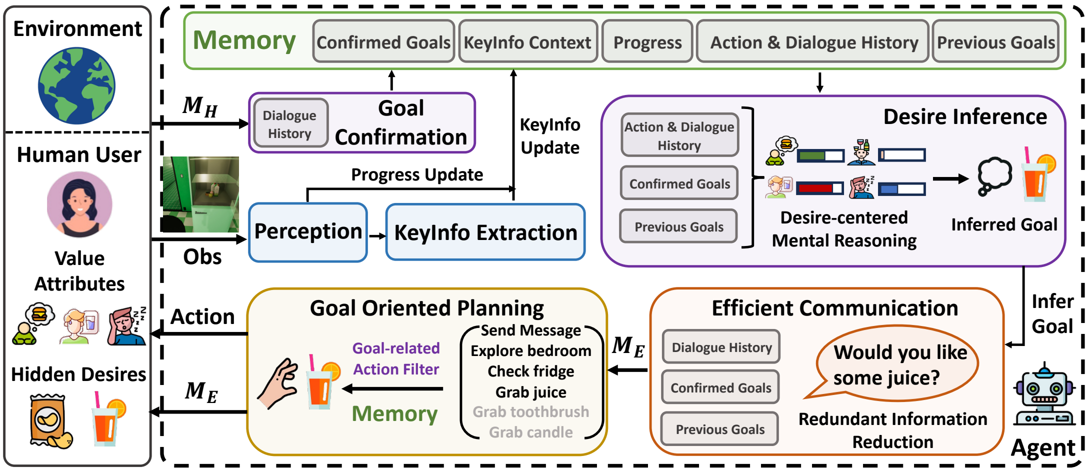

Figure 3: Overview of the FAMER framework. FAMER comprises three key components: KeyInfo Extraction, Desire-Centered Mental Reasoning (including Goal Confirmation and Desire Inference), and Efficient Communication. These are supported by a memory module, perception module, and planning module, which together form an integrated pipeline for embodied agent-human adaptation.

the RGB-D data. These outputs are used to build a scene graph that encodes object locations, categories, and spatial relationships.

We then introduce the first core component of FAMER: Key Information Extraction. With the confirmed and inferred goals context, this LLM-based module filters and stores goal-relevant information extracted from the scene graph into a dedicated cross-episode memory buffer. For example, if the agent identifies that juice is located in the fridge within the kitchen, it stores the structured entry "juice in fridge in kitchen" in memory. During subsequent planning phases, this stored information is used to guide attention to known facts and reduce redundant exploration. As a result, the agent is better equipped to reuse previously acquired knowledge across episodes, improving task efficiency.

4.2 Desire-centered Mental Reasoning

This module is composed of two interconnected components: Goal Confirmation and Desire Inference, as illustrated in Figure 3. Together, they enable the agent to infer the user's underlying desires.

The Goal Confirmation component extracts confirmed goals from the user's responses by LLM reasoning. For example, if the agent asks, "Do you want some juice?" and the user replies, "Correct! And I also want something crunchy," the system confirms that juice is one of the user's desired items. This process grounds part of the goal set and reduces future uncertainty.

Following confirmation, the LLM-based Desire Inference component leverages the action & dialogue history, confirmed goals and past episode goals to reason about the user's mental state, including value attributes and desires. Since user values remain consistent across episodes, the agent can incrementally improve its inference accuracy over time. By maintaining an internal model of the user, the agent can focus on narrowing down the remaining potential goals and avoid repetitive or redundant guesses.

With both inferred and confirmed goals in hand, the agent filters out irrelevant actions during planning. As shown in Figure 3, if the current goals do not involve a toothbrush or candle, then actions involving those objects are ignored. This reduces distraction and helps the agent maintain focus on goal-relevant objects and activities, thereby improving planning efficiency and task performance.

4.3 Efficient Communication

Excessive or repetitive communication can diminish user satisfaction and hinder overall system efficiency. To address this, FAMER integrates an LLM-based Efficient Communication module that promotes purposeful, context-aware dialogue between the agent and the user.

This module leverages both the dialogue history and the agent's inferred model of the user's mental state to decide when and what to ask. Before initiating a new query, the agent performs an internal reflection over its current knowledge—what goals have been confirmed, which value attributes have been inferred, and what uncertainties remain.

This reflective mechanism helps avoid redundant or previously resolved questions, particularly across multi-episode interactions.

When communication is necessary, the agent formulates targeted, desire-aligned questions aimed at resolving specific ambiguities. For example, if the agent has inferred a preference for sweet items but is uncertain about texture, it may ask "I found a cupcake. Would you like it?" instead of issuing vague or open-ended queries. This focused interaction strategy minimizes the communication burden on the user while enabling the agent to efficiently acquire high-value information.

5 Experiment

5.1 Tasks & Metrics

We evaluate FAMER in two representative tasks instantiated within our proposed HA-Desire environment: **Prepare Afternoon Snack** and **Set Up Dinner Table**. Each task is associated with a five-dimensional value space that governs user preferences. The Snack task includes 10 potential goals, while the Table task contains 8. Each task is further divided into two levels: Medium and Large. In Medium setting, the number of target goals is 2, and the maximum episode length is 60 steps. In Large setting, the agent must satisfy 4 goals within 120 steps. For example, the Snack-M task involves a total of $C(10, 2) = 45$ possible goal combinations. The reason why we set the maximum goal space to 4 varying objects is that for daily tasks such as meal preparation, people often share a stable backbone of common objects and a small consideration subset of items (≈3–5) that vary depending on individual preferences due to limited cognitive constraints (Schank and Abelson, 1975; Wood et al., 2022; Hauser and Wernerfelt, 1990). Details of the tasks are provided in Appendix A.

For each task at a given level, every method is evaluated over 6 independent runs. In each run, the agent interacts with the human user for $K = 3$ episodes, with the user's value attributes fixed throughout the interaction. All reported results are averaged over the 6 runs. We evaluate performance using the following metrics:

Score: Given N total goals, the agent receives a reward of $\frac{1}{N}$ for each correct goal achieved. Completing an incorrect or distracting goal incurs a penalty of $-\frac{1}{2N}$. The maximum achievable score per episode is 1, corresponding to the completion of all goals without any mistakes.

Communication Cost: The total number of tokens exchanged in messages between the agent and the user, including both queries and responses.

Notably, HA-Desire supports diverse objects and scenes, and goal generation is automated, making it easy to extend to other tasks. Further details of the environments are provided in Appendix A.

5.2 Baselines & Ablations

We compare FAMER against three baselines and three ablated variants.

CoELA (Zhang et al., 2024b): An LLM-based multiagent cooperation framework. Original CoELA assumes full observability of goals. To adapt it to our setting, we modify its prompting so that the agent is only aware of the potential goal set and the number of target goals.

ProAgent (Zhang et al., 2024a): A proactive LLM-based agent framework designed for collaboration in fixed-goal tasks. It lacks mechanisms for handling goal uncertainty or communication. We extend ProAgent by adding cross-episode memory to support adaptation to latent user desires.

MHP: An MCTS-based Hierarchical goal-sampling Planner adapted from the Watch-and-Help Challenge (Puig et al., 2021). We introduce subgoal sampling to handle uncertain goals and maintain a cross-episode success memory. Like ProAgent, MHP does not support communication.

FAMER w/o Desire: Removes the Goal Confirmation, Desire Inference, and goal-related action filtering modules.

FAMER w/o EC: Disables the Efficient Communication module, leading to less targeted and potentially redundant dialogue.

FAMER w/o KeyInfo: Removes the Key Information Extraction module, block the agent from memory of known object-location pairs.

5.3 Experimental Results

We evaluate performance using the two metrics on both Snack and Table tasks at two difficulty levels: Medium and Large. Results for the Snack-M and Snack-L tasks are shown in Figure 4, while those for Table-M and Table-L are presented in Figure 5. For each method, the three adjacent bars represent performance across three consecutive episodes with the same user.

From the figures, it is evident that FAMER consistently outperforms all three baselines across all metrics. In terms of score, FAMER achieves the

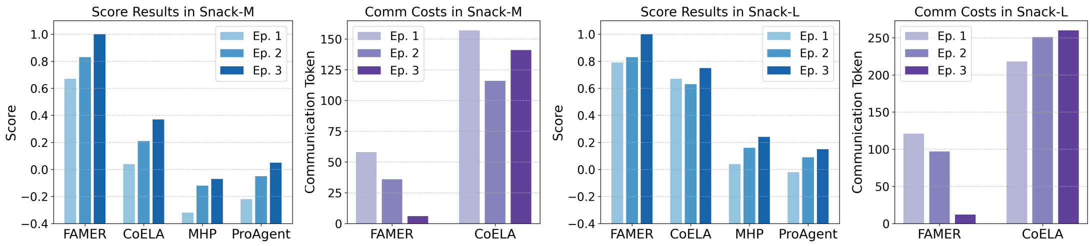

Figure 4: Quantitative results in Snack-M and Snack-L. Ep. denotes Episode, M denotes Medium, and L denotes Large, which applies to all subsequent figures and tables. FAMER achieves a perfect score by the third episode and outperforms baselines in both task score and communication efficiency.

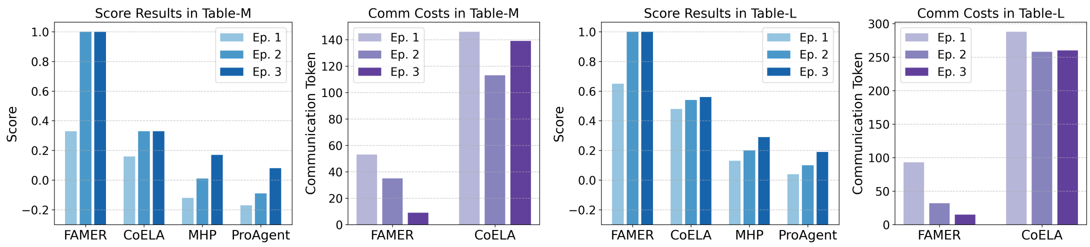

Figure 5: Quantitative results in Table-M and Table-L. FAMER achieves higher scores and superior communication efficiency compared to all baselines.

maximum value of 1.0 in the third episode, indicating that it successfully infers the desired goals of the human user within only 3 episodes and completes all of them. CoELA performs second best but falls short due to its reliance on long-context LLM prompting alone, which leads to occasional misinterpretation of user desires. This limitation will be further illustrated in the qualitative analysis in Appendix B.4. MHP and ProAgent perform the worst, as they lack communication capabilities and rely solely on trial-and-error to identify goals. Such an inefficient process often incurs penalties. Notably, their performance gradually improves across episodes, reflecting slow adaptation to latent user desires through repeated interactions.

In terms of communication efficiency, FAMER significantly outperforms CoELA, as shown in Figures 4 and 5. This efficiency stems from FAMER's reflection-based communication strategy, which avoids redundant questions. In contrast, CoELA frequently issues similar or vague queries due to its lack of explicit goal-tracking mechanisms.

5.4 Ablation Study

We further evaluate the contribution of each FAMER component through ablation studies on the Snack-M task. As shown in Figure 6, all three ablated variants exhibit performance degradation across the evaluated metrics. Among them, FAMER w/o KeyInfo is able to eventually achieve

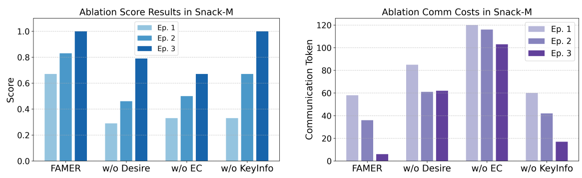

Figure 6: Ablation results on Snack-M. Removing any component of FAMER degrades performance in both scores and communication efficiency.

Snack-M (Evolve)	Ep.1	Ep.2	Ep.3	Ep.4	Ep.5	Ep.6
Score	0.64	0.77	0.85	0.61	0.71	0.79
Comm Cost	72	54	49	89	61	53

Table 1: Performance of FAMER on Snack-M under evolving user preferences. The proxy user's value attributes are resampled after Episode 3, requiring the agent to adapt to a new preference profile.

a full score in the third episode, similar to the full model, but its scores in the first two episodes are lower and it incurs slightly higher communication costs. This indicates that the Key Information Extraction module primarily improves efficiency by reducing unnecessary exploration.

In contrast, substantial performance drops are observed in FAMER w/o Desire and w/o EC, underscoring the central roles of desire modeling and efficient communication in agent-human adaptation. Without goal confirmation, desire inference, and goal-aligned action filtering, the agent strug-

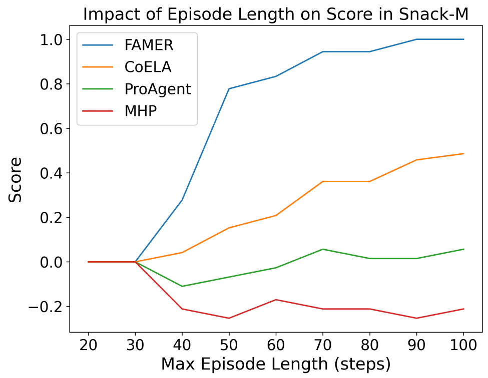

Figure 7: Impact of maximum episode length on average score in Snack-M.

Method	Ep. 1	Ep. 2	Ep. 3
CoELA	0.67	0.83	0.83
MHP	0.42	0.33	0.42
ProAgent	0.50	0.50	0.58
FAMER	0.67	1.00	1.00

Table 2: Score results across three episodes in the Human Reveal Goals setting on Snack-M.

Agent LLM	Human Proxy LLM			
	GPT-4o	GPT-5	Gemini2.5	Qwen3
GPT-4o	0.85	0.84	0.83	0.84
GPT-5	0.87	0.85	0.86	0.85
Gemini2.5	0.81	0.84	0.85	0.83
Qwen3	0.81	0.82	0.79	0.81

Table 3: Average score over three episodes on the Snack task using different LLM backbones for the ego agent and the proxy human user.

gles to correctly interpret user intent, leading to incorrect or inefficient actions. Moreover, the sharp increase in communication cost for FAMER w/o EC highlights that reflection-based communication is crucial for minimizing redundant messages and maintaining efficiency.

Evolving Desires. We further evaluate whether FAMER can handle changing user preferences on Snack-M over six episodes. In this setting, the proxy user's value attributes are fixed for Episodes 1-3, then resampled before Episode 4, yielding a new preference profile for Episodes 4-6. Meanwhile, the specific goal set is resampled in every episode conditioned on the current value attributes. Results in Table 1 show that during Episodes 1-3, FAMER steadily improves from 0.64 to 0.85 in score while reducing communication cost, indicating effective reuse of inferred user values to narrow the hypothesis space. Compared with the original setting where both values and goals remain fixed, performance is slightly lower due to per-episode goal variation. After the value shift at Episode 4, performance temporarily drops to 0.61 and communication cost rises to 89, reflecting the need to adapt to a now unfamiliar user profile. However, FAMER quickly recovers in subsequent episodes, reaching a score of 0.79 with a reduced communication cost by Episode 6. These results demonstrate that FAMER can effectively adapt to evolving user preferences through continued interaction.

Impact of Episode Length on Score. We evaluate how maximum episode length affects performance on Snack-M by varying the limit from 20 to 100 steps and averaging scores across three episodes (Figure 7). When the limit is below 30, all methods score zero, as no goal can be completed. With more steps, FAMER and CoELA improve in-episode due to their ability to communicate, while FAMER reaches a perfect score once the limit exceeds 90 steps, showing it can fully adapt in the first episode given enough steps. In contrast, ProAgent and MHP may decline as episode length grows, since they rely on trial-and-error for adaptation, which leads to more wrong-goal penalties.

Human Reveal Goals. We also test a setting where the human user directly reveals the goals in each episode, removing the need for goal inference. Results on Snack-M (Table 2) show all methods improve, but FAMER still leads. Its Key Information Extraction module stores goal-related details like object positions across episodes, avoiding redundant exploration, while baselines such as CoELA must still search extensively in each episode. Combined with goal-oriented planning that prunes irrelevant actions, this allows FAMER to outperform baselines even when goals are explicitly given.

Different LLMs. We further test robustness of our benchmark and framework by varying the LLMs used for both the agent and the proxy user on the Snack task. As shown in Table 3, the average score varies only slightly across different models. This indicates that our evaluation and method are robust to the choice of underlying LLMs.

5.5 Human Study

To evaluate how FAMER performs with real human users, we recruited 8 participants to serve as users in two scenarios: Snack-M and Table-M. Each participant was randomly assigned a set of value attributes and asked to act as the human user

Method	Satisfaction	Helpfulness	Comm Efficiency
CoELA	4.4±0.6	4.7±0.6	4.2±0.6
FAMER w/o Desire	4.4±0.5	4.6±0.6	4.8±0.5
FAMER w/o EC	4.8±0.7	5.2±0.7	4.3±0.5
FAMER	**5.6±0.6**	**5.9±0.6**	**5.6±0.6**

Table 4: Human study results on Snack-M and Table-L. Participants rated Satisfaction, Helpfulness, and Communication Efficiency on a 7-point Likert scale.

in HA-Desire, communicating with the agents and evaluating whether they successfully satisfied the assigned values.

For each task, participants interacted with four agents: CoELA, FAMER w/o Desire, FAMER w/o EC, and FAMER, in random order. Each agent was tested for 3 episodes, following the same protocol as in the simulation setting. After each session, participants rated the assigned agent on a 7-point Likert scale with respect to three dimensions: (1) Satisfaction: *I am satisfied with the overall performance of the agent*. (2) Helpfulness: *The agent helped me obtain what I wanted*. (3) Communication Efficiency: *The agent communicated efficiently without asking redundant or irrelevant questions*. Each participant evaluated all four agents across both tasks for three episodes, resulting in a total of 192 episodes of human-agent interactions.

Results are presented in Table 4. FAMER consistently achieved the highest ratings across all three criteria, further supporting its effectiveness in real human-agent adaptation scenarios.

6 Conclusion

This work investigates a central challenge in human–agent interaction: enabling embodied agents to quickly adapt to unfamiliar users and provide proactive assistance while minimizing communication effort. In real-world household settings, interaction is long-term and repeated, and users expect agents to gradually learn their preferences rather than relying on precise instructions. To support systematic study of this problem, we introduce HA-Desire, a 3D embodied simulation environment with value-driven proxy users and natural language interaction, designed to evaluate communication-efficient user adaptation in realistic scenarios.

Building on HA-Desire, we propose FAMER, a framework that enables fast and communication-efficient desire alignment through the integration of desire-centered reasoning, goal-relevant memory, and selective communication. By explicitly modeling user preferences from interaction and reducing redundant dialogue, FAMER supports proactive agent behavior in realistic tasks.

Experimental results across multiple household tasks demonstrate that FAMER consistently improves both task success and interaction efficiency compared to strong baselines, with ablation studies confirming the contribution of each component. Additional analysis, including the human-revealed-goal setting, the impact of episode length, different LLM backbones, and a human-subject study, further validates the robustness of our approach for proactive human-agent interaction.

Limitations

A primary limitation of this work is that our evaluation is conducted in a simulated environment with limited scenarios and does not include deployment on a real mobile robot interacting with human users in physical settings. While HA-Desire enables controlled and scalable study of communication-efficient embodied human-agent interaction, real-world deployment introduces additional challenges such as perception noise, action uncertainty, and richer social dynamics. Addressing these factors through real-robot experiments and in-the-wild human–agent interaction remains an important direction for future work. In addition, like other interactive AI systems, our approach may be vulnerable to adversarial or malicious attack, where manipulated interaction signals could cause a helper agent to behave in unintended or potentially harmful ways, highlighting the need for robustness and safety mechanisms in future work.

Acknowledgments

This research is supported by the National Science and Technology Major Project (2022ZD0114904), the National Natural Science Foundation of China (No. 62376006), Beijing Natural Science Foundation L252010, NSFC-62406010, and the Fundamental Research Funds for the Central Universities.

References

Stefano V Albrecht and Subramanian Ramamoorthy. 2015. A game-theoretic model and best-response learning method for ad hoc coordination in multiagent systems. *arXiv preprint arXiv:1506.01170*.

Nolan Bard, Jakob N Foerster, Sarath Chandar, Neil Burch, Marc Lanctot, H Francis Song, Emilio

Parisotto, Vincent Dumoulin, Subhodeep Moitra, Edward Hughes, and 1 others. 2020. The hanabi challenge: A new frontier for ai research. *Artificial Intelligence*, 280:103216.

James R Bettman, Mary Frances Luce, and John W Payne. 1998. Constructive consumer choice processes. *Journal of consumer research*, 25(3):187–217.

Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. 2019. On the utility of learning about humans for human-ai coordination. *Advances in Neural Information Processing Systems*, 32.

Shuo Chen, Ewa Andrejczuk, Zhiguang Cao, and Jie Zhang. 2020. Aateam: Achieving the ad hoc teamwork by employing the attention mechanism. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7095–7102.

Herbert H Clark and Susan E Brennan. 1991. Grounding in communication.

Brandon Cui, Hengyuan Hu, Luis Pineda, and Jakob Foerster. 2021. K-level reasoning for zero-shot coordination in hanabi. *Advances in Neural Information Processing Systems*, 34:8215–8228.

Josef Dai, Xuehai Pan, Ruiyang Sun, Jiaming Ji, Xinbo Xu, Mickel Liu, Yizhou Wang, and Yaodong Yang. 2023. Safe rlhf: Safe reinforcement learning from human feedback. *arXiv preprint arXiv:2310.12773*.

Weihua Du, Qiushi Lyu, Jiaming Shan, Zhenting Qi, Hongxin Zhang, Sunli Chen, Andi Peng, Tianmin Shu, Kwonjoon Lee, Behzad Dariush, and 1 others. 2024. Constrained human-ai cooperation: An inclusive embodied social intelligence challenge. In *Thirty-eighth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

Jaime F Fisac, Monica A Gates, Jessica B Hamrick, Chang Liu, Dylan Hadfield-Menell, Malayandi Palaniappan, Dhruv Malik, S Shankar Sastry, Thomas L Griffiths, and Anca D Dragan. 2020. Pragmatic-pedagogic value alignment. In *Robotics Research: the 18th International Symposium ISRR*, pages 49–57. Springer.

Hitesh Goel and Hao Zhu. 2025. Lifelong sotopia: Evaluating social intelligence of language agents over lifelong social interactions. *arXiv preprint arXiv:2506.12666*.

John R Hauser and Birger Wernerfelt. 1990. An evaluation cost model of consideration sets. *Journal of consumer research*, 16(4):393–408.

Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2961–2969.

Laura M Hiatt, Cody Narber, Esube Bekele, Sangeet S Khemlani, and J Gregory Trafton. 2017. Human modeling for human–robot collaboration. *The International Journal of Robotics Research*, 36(5-7):580–596.

Hengyuan Hu, Adam Lerer, Brandon Cui, Luis Pineda, Noam Brown, and Jakob Foerster. 2021. Off-belief learning. In *International Conference on Machine Learning*, pages 4369–4379. PMLR.

Hengyuan Hu, Adam Lerer, Alex Peysakhovich, and Jakob Foerster. 2020. "other-play" for zero-shot coordination. In *International Conference on Machine Learning*, pages 4399–4410. PMLR.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, and 1 others. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Jiaming Ji, Tianyi Qiu, Boyuan Chen, Borong Zhang, Hantao Lou, Kaile Wang, Yawen Duan, Zhonghao He, Jiayi Zhou, Zhaowei Zhang, and 1 others. 2023. Ai alignment: A comprehensive survey. *arXiv preprint arXiv:2310.19852*.

Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. 2023. Camel: Communicative agents for "mind" exploration of large language model society. *Advances in Neural Information Processing Systems*, 36:51991–52008.

Jie Liu, Pan Zhou, Yingjun Du, Ah-Hwee Tan, Cees GM Snoek, Jan-Jakob Sonke, and Efstratios Gavves. 2025. Capo: Cooperative plan optimization for efficient embodied multi-agent cooperation. In *International Conference on Learning Representations*.

Andrei Lupu, Brandon Cui, Hengyuan Hu, and Jakob Foerster. 2021. Trajectory diversity for zero-shot coordination. In *International Conference on Machine Learning*, pages 7204–7213. PMLR.

Long Ma, Yuanfei Wang, Fangwei Zhong, Song-Chun Zhu, and Yizhou Wang. 2024. Fast peer adaptation with context-aware exploration. In *International Conference on Machine Learning*.

Reuth Mirsky, William Macke, Andy Wang, Harel Yedidsion, and Peter Stone. 2020. A penny for your thoughts: The value of communication in ad hoc teamwork. *International Joint Conference on Artificial Intelligence*.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22.

Joon Sung Park, Carolyn Q Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S Bernstein. 2024. Generative agent simulations of 1,000 people. *arXiv preprint arXiv:2411.10109*.

Jinghua Piao, Yuwei Yan, Jun Zhang, Nian Li, Junbo Yan, Xiaochong Lan, Zhihong Lu, Zhiheng Zheng, Jing Yi Wang, Di Zhou, and 1 others. 2025. Agentsociety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society.

Xavier Puig, Kevin Ra, Marko Boben, Jiaman Li, Tingwu Wang, Sanja Fidler, and Antonio Torralba. 2018. Virtualhome: Simulating household activities via programs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8494–8502.

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

Neil Rabinowitz, Frank Perbet, Francis Song, Chiyuan Zhang, SM Ali Eslami, and Matthew Botvinick. 2018. Machine theory of mind. In *International Conference on Machine Learning*, pages 4218–4227. PMLR.

Muhammad A Rahman, Niklas Hopner, Filippos Christianos, and Stefano V Albrecht. 2021. Towards open ad hoc teamwork using graph-based policy learning. In *International Conference on Machine Learning*, pages 8776–8786. PMLR.

Manish Chandra Reddy Ravula. 2019. Ad-hoc teamwork with behavior-switching agents. In *International Joint Conferences on Artificial Intelligence*.

Roger C Schank and Robert P Abelson. 1975. *Scripts, plans, and knowledge*. Yale University New Haven, CT.

Peter Stone, Gal Kaminka, Sarit Kraus, and Jeffrey Rosenschein. 2010. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 24, pages 1504–1509.

DJ Strouse, Kevin McKee, Matt Botvinick, Edward Hughes, and Richard Everett. 2021. Collaborating with humans without human data. *Advances in Neural Information Processing Systems*, 34:14502–14515.

Yiding Wang, Yuxuan Chen, Fangwei Zhong, Long Ma, and Yizhou Wang. 2025. Simulating human-like daily activities with desire-driven autonomy. In *International Conference on Learning Representations*.

Yuanfei Wang, Fangwei Zhong, Jing Xu, and Yizhou Wang. 2022. Tom2c: Target-oriented multi-agent communication and cooperation with theory of mind. In *International Conference on Learning Representations*.

Wendy Wood, Asaf Mazar, and David T Neal. 2022. Habits and goals in human behavior: Separate but interacting systems. *Perspectives on Psychological Science*, 17(2):590–605.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations (ICLR)*.

Luyao Yuan, Xiaofeng Gao, Zilong Zheng, Mark Edmonds, Ying Nian Wu, Federico Rossano, Hongjing Lu, Yixin Zhu, and Song-Chun Zhu. 2022. In situ bidirectional human-robot value alignment. *Science Robotics*, 7(68):eabm4183.

Ceyao Zhang, Kaijie Yang, Siyi Hu, Zihao Wang, Guanghe Li, Yihang Sun, Cheng Zhang, Zhaowei Zhang, Anji Liu, Song-Chun Zhu, and 1 others. 2024a. Proagent: building proactive cooperative agents with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17591–17599.

Hongxin Zhang, Weihua Du, Jiaming Shan, Qinhong Zhou, Yilun Du, Joshua B Tenenbaum, Tianmin Shu, and Chuang Gan. 2024b. Building cooperative embodied agents modularly with large language models. In *International Conference on Learning Representations*.

Xuhui Zhou, Hao Zhu, Leena Mathur, Ruohong Zhang, Haofei Yu, Zhengyang Qi, Louis-Philippe Morency, Yonatan Bisk, Daniel Fried, Graham Neubig, and 1 others. 2023. Sotopia: Interactive evaluation for social intelligence in language agents. *arXiv preprint arXiv:2310.11667*.

A Environment Details

HA-Desire is built upon VirtualHome (Puig et al., 2018), a widely adopted testbed for embodied multi-agent cooperation. In this work, we modify the simulation to specifically address the challenge of embodied human-agent interaction with a focus on desire alignment. In the following sections, we provide further details on the human user agent and the tasks designed to evaluate this problem.

A.1 Human User

As discussed in Section 3, HA-Desire is designed to evaluate the ability of embodied agents to rapidly infer the underlying desires of human users and take actions to fulfill those desires. To achieve this, we integrate a proxy human user within the environment, which serves two main functions:

1. The human user determines the specific goal set from a set of potential goals, based on a vague task description and a sampled set of value attributes.

2. The human user responds to the agent's inquiries in natural language, ensuring that the goal set is not directly revealed. Instead, the user implies their goals by providing hints about object properties, reflecting their underlying preferences.

To achieve these functions, we empower the human user with large language models (LLMs), specifically GPT-4o. The prompts used to generate goals and responses for communication are outlined below.

We incorporate a Chain of Thought prompt at the end of each query to encourage the human user to think more thoroughly, leading to more accurate goal selection and communication. After the LLM generates results, we instruct it to extract the exact goals or messages, forming the final output. Several variables are included in the prompts, prefixed with a "$". During inference, these variables are replaced with context-specific information.

Goal Generation:

```
I am Bob, a human user living at home
with a humanoid assistant named Alice. I
 have several personal value attributes
(e.g., hungry, thirsty, alcoholic), each
 rated at one of three levels: Not,
Somewhat, or Very. Given a specific set
of my current attribute states, along
with a high-level task description, your
 task is to select the most appropriate
goal set from a given potential set of
goals. For example, if I am Very thirsty
, the goal set should include beverages.
 If I am not SweetTooth, the goal set
should include less sweet food objects.
Please help me choose the best goal set
to reflect my value attributes. Select
$GOAL_CNT$ objects as the goal set.
Provide your answer as a comma-separated
 list of object names.
An example output is: cupcake, milk,
pudding.
Value Attribute: $Value$
Task: $Task$
Potential Goals: $GOAL$
Answer:
Let's think step by step.
```

Communication:

```
I am Bob, a human user living at home
with a humanoid assistant named Alice. I
 need Alice to help me with a household
task, which is described in a high-level
 instruction without specific goals
provided. I have several personal value
attributes (e.g., hungry, thirsty,
alcoholic) that determine the goal, but
this goal is only visible to me and not
to Alice. Since Alice is unaware of the
specific goal, she may ask me questions
about it. However, I do not want to
directly tell her the goal; instead, I
want her to gradually understand my
preferences and needs through
interaction. Over time, I expect her to
infer the goal on her own without
needing to ask. The following Status
shows the number of EPISODEs I have
interacted with Alice. The larger the
number is, the less willing I am to talk
 to Alice. If Alice proposes a goal or
action that is incorrect, I can point
out the mistake. If the dialogue
progresses but the task is not
progressing, I may be more inclined to
correct her by hinting at one of the
goals, but I will never reveal the
entire goal set at once unless Alice
herself proposes the exact whole goal
set.
Task: $Task$
Status: This is the $EPISODE$-th time I
interact with Alice.
Goal: $GOAL$
Progress: $PROGRESS$
Alice Previous Action: $ACTION_HISTORY$
Previous Dialogue History:
Alice: "Hi, I'll let you know if I find
any goal objects and finish any subgoals
, and ask for your instruction and
clarification when necessary."
Bob: "Thanks! Let me know if you are
uncertain about the goal objects."
$DIALOGUE_HISTORY$
Alice asks this time: $QUESTION$
Note:
1. The generated message should be
accurate and brief. Use simple
expressions more often. Do not generate
repetitive messages.
2. Do not directly tell Alice the
specific goal name at the first time. (
The most important). Instead, hint
through some vague descriptions that
reveal some properties of the goals,
such as sweet, crunchy, alcoholic, etc.
3. Confirm Alice's correct guess (or
partly correct guess). But if the guess
contains too many objects compared to
the goal set, I should not confirm any
of the objects and hint at some
descriptions instead. For example,
suppose the correct goal set is [apple,
orange]. If Alice guesses [bottle,
banana, apple, orange, milk, chips] in
one round, I should not confirm the goal
 as the guess contains too many objects.
 If Alice guesses [chips, apple], then I
 should confirm that apple is correct,
even though the chips guess is wrong. If
 Alice guesses [apple, orange], I should
 say these two objects are exactly what
I want.
4. Do NOT guess the location of objects
```

```
or tell Alice where to find the goal
objects.
5. Be aware of the number of EPISODEs: A
 larger number means lower communication
 willingness.
6. Even if the guess is incomplete, I
should confirm the correct goals if
there are any.
7. If Alice has guessed something
correctly in previous dialogue, try to
focus on the new goal objects in this
message.
Please think step by step:
```

A.2 Tasks

As described in Section 5, we evaluate our method, along with baselines and ablations, on two tasks: Prepare Afternoon Snack and Set Up Dinner Table. The details of these tasks are provided below.

Snack. This task involves ten potential goals: cupcake, wine, milk, cereal, chips, apple, juice, pudding, creamybuns, and chocolatesyrup. The value space consists of five dimensions: Hungry, Thirsty, SweetTooth, Fruitarian, and Alcoholic, each of which takes one of three discrete levels: Not, Somewhat, or Very. The human user randomly samples values for these dimensions and then uses a language model to generate the corresponding goal set. The Snack-M level represents the medium difficulty, with 2 goals and a maximum of 60 steps. The Snack-L level represents the large difficulty, with 4 goals and a maximum of 120 steps. Performance is evaluated using two metrics: score and communication cost, as described in Section 5.

Table. This task includes eight potential goals: coffeepot, breadslice, cutleryknife, mug, plate, wineglass, cutleryfork, and waterglass. The value space consists of five dimensions: NeedRefresh, Thirsty, MeatLove, CaffeinTolerable, and Alcoholic. The value levels, number of goals, step limit, and evaluation metrics are identical to those in the Snack task.

The Snack and Table tasks serve as representative home assistance tasks in our evaluation. However, the VirtualHome simulator supports a wide range of object assets and activities, allowing for the easy extension of HA-Desire to additional household tasks. By defining the appropriate value space and potential goal set, new tasks can be seamlessly incorporated into the environment.

B Experiment Details

B.1 Computing Resource

The experiments were conducted on a workstation equipped with an NVIDIA GeForce RTX 4090 GPU and an Intel Core i9-13900K CPU. The large vision-language model used in this study is GPT-4o.

B.2 FAMER Implementation

The perception module of FAMER in HA-Desire follows the design of CoELA (Zhang et al., 2024b). It employs a Mask-RCNN to generate segmentation masks from RGB images, then combines them with depth information to build 3D point clouds of objects. From these, the agent extracts high-level information such as the position of key objects and constructs a structured semantic map for downstream reasoning and planning.

The memory module of FAMER maintains several types of information, as illustrated in Figure 3: Confirmed Goals, KeyInfo Context, Task Progress, Previously Achieved Goals, and Action & Dialogue History. The first four categories represent compact summaries of task state and goal inference, and thus grow slowly during interaction; each typically contains fewer than 20 entries, making it feasible to store them entirely. In contrast, Action & Dialogue History grows linearly with the number of steps. To manage this, we retain only the latest 10 entries in the memory context. Since the key information context preserves important earlier details, essential knowledge from prior interactions is not lost.

The key components of FAMER, including KeyInfo Extraction, Goal Confirmation, Desire Inference, Efficient Communication, and Goal-oriented Planning, are all LLM-driven modules. These modules take related contexts as input with instruction templates for an LLM, which reason about the required information.

B.3 Baselines

Here, we provide the detailed prompts for CoELA and ProAgent, which are adapted from their original versions to help the agents account for uncertain goals.

CoELA Planning

```
I'm $AGENT_NAME$, a humanoid home
assistant. I'm in a hurry to finish the
housework for my owner $OPPO_NAME$. I
know the high-level instruction of the
```

```
task, but I am not certain about the
specific goal determined by $OPPO_NAME$.
Given the potential goal, dialogue
history, and my progress and previous
actions, please help me infer and choose
 the best available action to achieve
the underlying goal as soon as possible.
Note that I can hold two objects at a
time and there are no costs for holding
objects. All objects are denoted as <
name> (id), such as <table> (712).
Task Name: $Task$
Potential Goal: $GOAL_CNT$ object(s)
determined by human user from the set
$GOAL$. Put them $REL_TARGET$
Progress: $PROGRESS$
Dialogue history:
Alice: ""Hi, I'll let you know if I find
 any goal objects and finish any
subgoals, and ask for your instruction
and clarification when necessary.""
Bob: ""Thanks! Let me know if you are
uncertain about the goal objects.""
$DIALOGUE_HISTORY$
Previous actions: $ACTION_HISTORY$
Available actions:
$AVAILABLE_ACTIONS$
Answer:
```

CoELA Communication

```
I'm $AGENT_NAME$, a humanoid home
assistant. I'm in a hurry to finish the
housework for my owner $OPPO_NAME$. I
know the high-level instruction of the
task, but I am not certain about the
specific goal determined by $OPPO_NAME$.
Given the potential goal, dialogue
history, and my progress and previous
actions, please help me generate a short
 message to send to my owner $OPPO_NAME$
 to help us achieve the underlying goal
as soon as possible.
Note that I can hold two objects at a
time and there are no costs for holding
objects. All objects are denoted as <
name> (id), such as <table> (712).
Potential Goal: $GOAL_CNT$ objects
determined by human user from the set
$GOAL$. Put them $REL_TARGET$
Progress: $PROGRESS$
Previous actions: $ACTION_HISTORY$
Dialogue history:
Alice: ""Hi, I'll let you know if I find
 any goal objects and finish any
subgoals, and ask for your instruction
and clarification when necessary.""
Bob: ""Thanks! Let me know if you are
uncertain about the goal objects.""
$DIALOGUE_HISTORY$

Note: The generated message should be
accurate and brief. Do not generate
repetitive messages.
```

ProAgent

```
I'm $AGENT_NAME$, a humanoid home
assistant. I'm in a hurry to finish the
housework for my owner $OPPO_NAME$. I
know the high-level instruction of the
task, but I am not certain about the
specific goal determined by $OPPO_NAME$.
Given the potential goal, my progress,
and previous actions, please help me
infer and choose the best available
action to achieve the underlying goal as
 soon as possible.
Note that I can hold two objects at a
time and there are no costs for holding
objects. All objects are denoted as <
name> (id), such as <table> (712).

Potential Goal: $GOAL_CNT$ object(s)
determined by human user from the set
$GOAL$. Put them $REL_TARGET$.

Important Instruction:
$AGENT_NAME$ has previously achieved and
 found these subgoals.This is its
success experience.
You should focus only on actions that
help achieve the goal items, i.e., those
 in the target set provided.
Ignore or deprioritize any actions
unrelated to acquiring or placing goal
items.
When reviewing previous successful
experiences, only reuse or adapt steps
that directly contribute to acquiring or
 placing goal items.For example, ignore
exploration, object grabbing, or placing
 steps for items not included in the
current potential goal set.
When the current situation even
partially matches any past success (e.g
., similar object types, room layout, or
 goal structure), you should prioritize
reusing or adapting the proven action
sequences:
$HISTORY_OF_SUCCESSFUL_SUBGOALS$

Progress: $PROGRESS$

Previous actions: $ACTION_HISTORY$

Belief State: $BELIEF_STATE$

Available actions:
$AVAILABLE_ACTIONS$

Required Output Format:
- Analysis: [Infer and choose the best
available action to achieve the
underlying goal]
- Best Next Action: [Single most optimal
 action from available options]
- Intention for $OPPO_NAME$'s Underlying
 Goal: [Inference about the true goal]
```

B.4 Qualitative Analysis

To further highlight FAMER's strengths, we present an intuitive comparison against CoELA on the Snack-L task. As illustrated in Figure 8, we visualize the agents' behavior through a series of key frames sampled across the episode. In this

Figure 8: Qualitative comparison between CoELA and FAMER on the Snack-L task. The figure shows a sequence of key frames illustrating agent behavior across one episode. CoELA exhibits three typical failure modes: 1. Misinterpreting latent user desire 2. Redundant actions 3. Excessive, repetitive communication. FAMER demonstrates more accurate desire inference, targeted questions, and efficient planning. It successfully identifies all four goals with minimal trial-and-error and completes the task with fewer steps and lower communication cost.

example, Alice refers to the ego agent and Bob refers to the human user. During task execution, the CoELA agent demonstrates three typical issues that contribute to its inferior performance.

First, CoELA struggles to correctly extract and infer desires. For instance, when the user says, "I want something crunchy or refreshing," which aligns with chips and juice, CoELA incorrectly interprets this as a preference for apple, and retrieves it as the first item. Similarly, in step 6, when the user mentions wanting "something that complements tea," the agent mistakenly infers cupcake instead of the intended milk. These errors illustrate CoELA's limited ability to perform precise desire inference, particularly in the face of ambiguous or indirect language.

Second, CoELA exhibits repeated and inconsistent behavior due to insufficient integration between planning and memory. In steps 3 and 4, the agent redundantly grabs and places an apple on the coffee table, mistakenly treating it as an unfulfilled goal. This reflects a lack of attention to confirmed goals or past actions. In contrast, FAMER incorporates goal-aligned action filtering to suppress such irrelevant behaviors once a goal has been ruled out.

Third, CoELA engages in redundant communication. As shown in Figure 8, the agent repeatedly mentions creamybuns and chips to the user, even after those items have already been retrieved and confirmed. This not only wastes communication bandwidth but also reflects poor tracking of dialogue.

In contrast, the FAMER agent asks focused questions to resolve uncertainty. Within a limited number of interactions, it successfully infers all four desired items and efficiently retrieves and places them on the coffee table. This example illustrates FAMER's advantages in goal inferring, memory-informed planning, and communication efficiency, enabling superior performance in complex scenarios.

C Ethics statement

This work includes a human-subject study designed to evaluate human-agent interaction in household tasks. The study involved 8 adult participants, all of whom were volunteers recruited from the authors' institution. Prior to participation, individuals were informed of the study objectives, the nature of the tasks, and the type of data to be collected. No personally identifiable information was collected, and all participants were compensated according to standard practices. All procedures were conducted in accordance with ethical standards. The tasks posed no physical or psychological risks to participants, as interactions were limited to computer-based simulations and short surveys.

D AI Assistant Usage

Besides the LLM-driven proxy user construction in HA-Desire, Large Language Models were utilized exclusively for language refinement, specifically to improve wording and to identify grammatical and spelling errors. The content and intellectual contributions of this manuscript were generated entirely by the authors. LLMs were employed to enhance clarity and readability of the text.