

MEMORY-AUGMENTED LARGE LANGUAGE MODEL-BASED AGENT WITH CROSS-TASK EXPERIENCE LEARNING

Anonymous authors

Paper under double-blind review

ABSTRACT

Large Language Model (LLM)-based agents have demonstrated impressive capabilities in complex decision-making and multi-turn instruction-following tasks. To enhance knowledge retention and contextual adaptability, recent work has equipped these agents with memory modules that store and reuse historical interaction experiences. However, existing memory-augmented approaches face two key limitations: they often require large amounts of interaction data during early training to reach competitive performance, resulting in low data efficiency; and they rely on static, self-derived experience reuse strategies, limiting their ability to adapt when prior learning is insufficient and preventing the use of transferable knowledge from related tasks. Building on these observations, in this paper, we propose a memory-augmented LLM agent with cross-task experience learning, designed to improve data efficiency and adaptability. Our method augments the conventional task-specific memory with an additional source experience memory that retains transferable knowledge from related but distinct tasks. We further introduce a dynamic memory retrieval mechanism that adaptively draws from both task and source memories, allowing the agent to balance prior task-specific experiences with cross-task knowledge according to the current context and progression. We validate the proposed method on the WebShop benchmark, which comprises diverse, multi-turn instruction-following tasks across product domains with varying semantic complexity. Experimental results show that our approach consistently outperforms state-of-the-art memory-augmented LLM agents in task success rate and generalization, demonstrating the effectiveness of the proposed memory architecture and retrieval mechanism.

1 INTRODUCTION

With the advancement of Transformer architectures and large-scale pre-training techniques, large language models (LLMs) have demonstrated remarkable capabilities in natural language understanding Shi et al. (2024), reasoning Huang & Chang (2023); Singh et al. (2023), and generation Stiennon et al. (2020); Bian et al. (2024); Akyürek et al. (2023). Building upon these strengths, LLMs are increasingly established as a core component of autonomous agents, with a growing number of studies leveraging them as the foundation for high-level decision-making Chen et al. (2019); Shinn et al. (2023); Shen et al. (2023); Zhu et al. (2023b); Sclar et al. (2023). These agents have demonstrated promising performance across diverse tasks such as decision planning Karpas et al. (2022); Du et al. (2023), information retrieval Nakano et al. (2022); Ruan et al. (2023), and tool use Schick et al. (2023). However, most existing LLM-based agents treat each interaction as an isolated event, lacking effective mechanisms to incrementally accumulate and reuse task-specific knowledge. This limitation severely impairs their ability to adapt strategies based on prior experience, reducing data efficiency and adaptability in continuously evolving task environments.

To address the limitation of LLM-based agents in retaining and leveraging prior experience for strategy adaptation, recent research has increasingly explored augmenting LLM-based agents with memory modules designed to support knowledge retention and contextual adaptation throughout multi-turn interactions Rana et al. (2023); Zhu et al. (2023a); Wang et al. (2025); Huang et al. (2023b). By incorporating mechanisms that emulate human memory processes, these approaches

054 aim to facilitate the gradual accumulation of task-specific knowledge and enhance the agent’s ability
055 to generalize across dynamic and evolving task environments. For instance, Reflective Linguistic
056 Programming (RLP) embeds the agent’s recent observations into prompts to simulate short-term
057 memory Fischer (2023). Generative agents extend this idea by combining short-term and long-term
058 memory structures Park et al. (2023). Reflexion introduces feedback-driven self-reflection that inte-
059 grates memory for iterative strategy refinement Shinn et al. (2023). More recently, REMEMBERER
060 frames the LLM as a reinforcement learning agent and augments it with an experience memory mod-
061 ule, representing a more advanced approach for retaining and retrieving interaction histories Zhang
062 et al. (2023). While memory-augmented designs improve adaptability, they still face two fundamen-
063 tal limitations. First, they typically rely solely on self-derived experience, lacking mechanisms to
064 leverage transferable knowledge from related tasks. As a result, they require extensive interaction
065 data in early training to compensate for knowledge gaps, leading to low data efficiency. Second,
066 most existing methods adopt static, manually designed memory retrieval strategies, which constrain
067 the agent’s ability to adaptively reuse experiences based on task context and progression.

068 Building on these observations, in this paper, we propose a memory-augmented LLM-based agent
069 with cross-task experience learning, aimed at improving both data efficiency and adaptability. In par-
070 ticular, our method augments the conventional task-specific memory with an additional source ex-
071 perience memory that retains transferable knowledge from related but distinct tasks, enabling more
072 efficient reuse of prior learning. Meanwhile, the task-specific memory itself is enhanced to more
073 effectively capture context-dependent information during ongoing interactions, thereby supporting
074 fine-grained adaptation to the current task. To effectively leverage these two forms of memory,
075 we further introduce a dynamic retrieval mechanism that adaptively balances task-specific experi-
076 ences with cross-task knowledge according to the current interaction state and progression, thereby
077 enhancing both decision quality and generalization across tasks. We evaluate our method on the
078 WebShop benchmark Yao et al. (2022), which involves diverse, multi-turn instruction-following
079 tasks spanning product domains with varying semantic contexts. Experimental results show that
080 our approach consistently achieves higher task success rates and better generalization compared to
081 state-of-the-art memory-augmented LLM-based agents, validating the effectiveness of the proposed
memory architecture and dynamic retrieval mechanism.

082 Our main contributions are summarized as follows: (1) We propose a memory-augmented LLM-
083 based agent with cross-task experience learning, which improves data efficiency and adaptability
084 by enabling the reuse of transferable knowledge from related but distinct tasks; (2) We design a
085 dynamic memory retrieval mechanism that adaptively balances task-specific and cross-task experi-
086 ences according to the agent’s interaction state and task progression, thereby supporting more effec-
087 tive decision-making; (3) We conduct extensive experiments on the WebShop benchmark, demon-
088 strating that our method consistently outperforms state-of-the-art memory-augmented LLM-based
089 agents in both task success rate and generalization across diverse task domains.

091 2 RELATED WORK

092
093 This section reviews the studies relevant to this work, including LLM-based agents and memory-
094 augmented LLM-based agents.

095
096 **LLM-Based Agents** Large language models (LLMs) have shown strong abilities in language un-
097 derstanding, reasoning, and decision-making, enabling their integration as core components in au-
098 tonomous agents for complex environments. Early approaches primarily relied on prompting for
099 action generation. For instance, SayCan Ahn et al. (2022) couples LLMs with an affordance model
100 for robotic control but lacks explicit feedback and memory for long-term planning. Inner Mono-
101 logue Huang et al. (2023a) incorporates environment-informed feedback for iterative refinement,
102 yet depends on predefined signals without mechanisms to accumulate knowledge. WebGPT Nakano
103 et al. (2022) enables browsing and information retrieval, but relies heavily on costly human feedback
104 and lacks explicit reasoning modeling. ReAct Yao et al. (2023) interleaves reasoning and action to
105 produce interpretable traces, improving task performance but still treating each interaction as iso-
106 lated. Despite these advances, most LLM-based agents remain limited in long-term, consistent, and
107 adaptive decision-making. By failing to incrementally accumulate and reuse prior experiences, they
struggle to adapt strategies over time, resulting in low data efficiency and brittle performance in
dynamic, multi-turn environments.

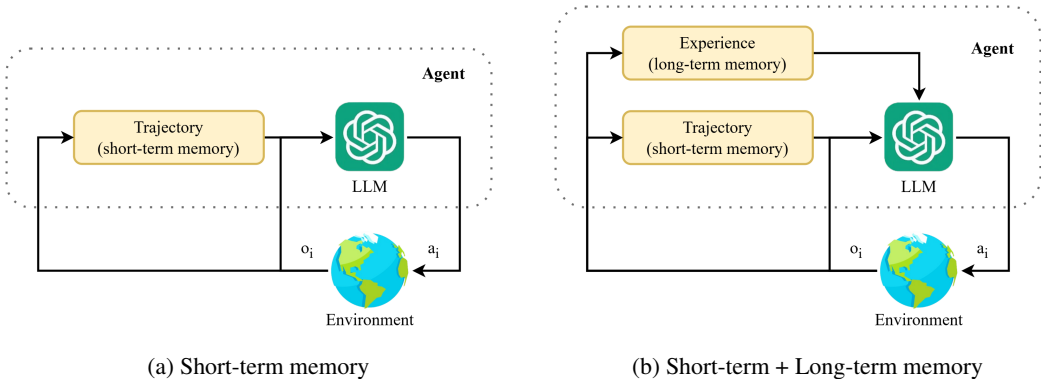


Figure 1: Architectures of LLM-based agents with different memory mechanisms.

Memory-augmented LLM-based Agents To address the limited capacity of LLM-based agents to accumulate and reuse experience, recent works introduce memory modules for knowledge retention and contextual adaptation across interactions. As shown in Fig. 1a, early efforts focus on short-term memory, leveraging only the current trajectory. For example, Reflective Linguistic Programming (RLP) Fischer (2023) injects recent observations directly into prompts, but lacks mechanisms to summarize or reuse past episodes, limiting long-horizon planning. DEPS Wang et al. (2023) and Corrective Re-Prompting Raman et al. (2022) utilize single-step corrective signals to refine responses. While effective at resolving local errors, they cannot accumulate or generalize knowledge across tasks. Consequently, such approaches treat each episode as isolated, preventing systematic knowledge reuse. To overcome these issues, more recent work explores long-term memory in combination with short-term modules (Fig. 1b). ChatDB Hu et al. (2023) and Ret-LLM Modarressi et al. (2024) employ external storage to persist environment states or feedback, but these modules are loosely coupled with task reasoning and updated infrequently, hindering responsiveness. Reflexion Shinn et al. (2023) integrates short- and long-term memory with feedback-driven self-reflection, yet its long-term store is severely capacity-limited, retaining only a small subset of experiences. GITM Zhu et al. (2023b) buffers successful trajectories, but its unstructured memory format impedes efficient retrieval and abstraction, restricting scalability. REMEMBERER Zhang et al. (2023) augments reinforcement learning with extended long-term memory for adaptive decision-making, but requires extensive interaction data and relies on fixed retrieval rules, limiting adaptability under shifting task demands.

Although memory-augmented agents have advanced beyond purely prompt-based approaches, existing methods still face critical limitations. First, they typically require large amounts of interaction data during early training to populate memory, resulting in low data efficiency. Second, most approaches rely on static, manually designed memory retrieval strategies, which restrict the flexibility and responsiveness of memory utilization. Consequently, these methods primarily exploit an agent’s own accumulated experiences, overlooking the potential benefits of cross-task knowledge transfer. Such transfer, however, holds considerable promise for improving data efficiency, particularly when the current task bears significant similarity to previously encountered ones.

3 PROPOSED METHOD

In this section, we present a detailed description of the proposed memory-augmented LLM-based agent with cross-task experience learning. We first outline the overall architecture, then elaborate on the target memory update process, and finally describe the dynamic retrieval mechanism.

3.1 ALGORITHM OVERVIEW

As illustrated in Figure 2, we propose a memory-augmented framework for LLM-based agents with cross-task experience learning. Our method extends conventional task-specific memory with an additional *source experience memory* that retains transferable knowledge from related but distinct

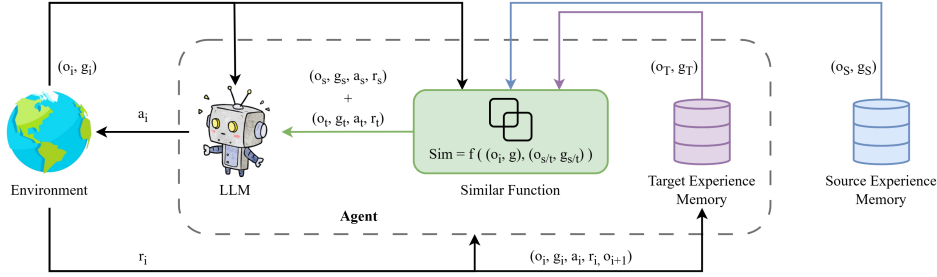


Figure 2: Architecture of the proposed memory-augmented LLM agent with cross-task experience learning

tasks. The agent dynamically leverages both memories to enhance decision-making across diverse tasks.

The *target experience memory* \mathcal{M}_t incrementally accumulates task-specific experiences during on-going interactions, enabling continual adaptation to the current task. Specifically, it stores experience tuples (o_t, g_t, a_t, r_t) , where o_t is the current observation, g_t the task goal, a_t the action taken, and r_t the resulting reward. The *source experience memory* \mathcal{M}_s retains high-utility experiences from previously completed tasks, facilitating cross-task generalization and reducing the need for extensive task-specific exploration. Each entry in \mathcal{M}_s is an experience tuple (o_s, g_s, a_s, r_s) , where o_s is a past observation, g_s the associated goal, a_s the action executed in that context, and r_s the reward received. Only successful or high-value experiences are included to minimize noise and maximize transfer effectiveness.

At each decision step i , the agent observes the current state and task goal (o_i, g_i) and retrieves relevant experiences from both \mathcal{M}_s and \mathcal{M}_t using a similarity-based retrieval function. Retrieval is dynamically weighted: early in an episode, both memories contribute to leverage transferable knowledge, while as \mathcal{M}_t accumulates sufficient experiences, retrieval gradually prioritizes target memory to mitigate negative transfer. The number of retrieved entries is adaptively adjusted to fit within the LLM’s token budget, ensuring maximal informativeness.

Retrieved experiences are converted into a structured natural language template capturing observation, goal, action, and reward (Figure 3). These are concatenated with the current observation and goal (o_i, g_i) to form a context-enriched prompt for the LLM, which then generates the next action a_i . After execution, the resulting transition $(o_i, g_i, a_i, r_i, o_{i+1})$ is appended to \mathcal{M}_t , enabling continual task-specific adaptation. Optionally, Q-values for state-action pairs are maintained to support similarity-weighted retrieval and ongoing policy refinement.

By structuring the agent around source and target memories with a dynamic retrieval mechanism, the proposed method efficiently leverages both transferable and task-specific experiences. This design improves decision quality, reduces the need for extensive task-specific interactions, and enhances generalization across diverse multi-turn tasks.

3.2 SOURCE MEMORY CONSTRUCTION AND TARGET EXPERIENCE UPDATE

To support cross-task knowledge transfer, we pre-construct a *source experience memory* \mathcal{M}_s that encapsulates transferable knowledge distilled from previously completed tasks. Each source experience is represented as a tuple

$$(o_s, g_s, a_s, r_s),$$

where o_s denotes the environment observation, g_s the task instruction or goal, a_s the executed action, and r_s the received reward. These experiences are collected from diverse source tasks to capture general strategies and patterns that are likely to benefit future tasks. Unlike the target memory, \mathcal{M}_s is *static* during task execution—it is not updated with new interactions—providing a stable prior that informs the agent’s decisions and facilitates cross-task generalization. The experiences in \mathcal{M}_s are encoded in a structured, language-readable format compatible with prompt construction, allowing them to be retrieved efficiently by the dynamic memory retrieval mechanism described in Section 3.3. This design ensures that the agent can leverage prior knowledge from related tasks,

```

216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234

```

Last 5 Actions:

- search[high speed hdmi cables 10 feet 5 pack] ...
- click[b08qsnm69h] b08qsnm69h is a 10 feet ...
- click[5-pack] The item meets most of the ...

Observation:

You have clicked 5-pack...

Instruction: i need high speed hdmi cables that ...

[Back to Search]

size: [1-pack][3-pack][5-pack][10-pack] ...

color: [black braided][blue braided] ...

4K HDMI Cables, Short Cord, 1ft [3 Pack] Braided ...

Price: \$15.xx

[Features]

...

[Buy Now]

Encouraged:

click[Buy Now] -> 1.0 The item meets all the requirements and is within the desired price range. You can proceed to purchase it.

Discouraged:

click[Back to Search] -> 0.0 The current item offers the desired options and I don't need to search for other items.

Figure 3: An example of the memory format used in the WebShop Yao et al. (2022) task. Each retrieved experience tuple $(o_{s/t}, g_{s/t}, a_{s/t}, r_{s/t})$ is transformed into a structured textual representation using a predefined template.

reducing the need for extensive exploration in the target task and improving data efficiency from the outset.

To enable continual adaptation within the current task, the *target experience memory* \mathcal{M}_t is incrementally updated with new trajectories collected during interaction. In contrast, the *source memory* \mathcal{M}_s remains fixed across tasks, serving as a stable prior distilled from previously completed tasks. This separation ensures that \mathcal{M}_s encodes transferable knowledge, while \mathcal{M}_t focuses on capturing task-specific regularities.

At each decision step i , given the environment state o_i and task instruction g_i , the agent executes an action a_i , receives a reward r_i , and transitions to o_{i+1} . The transition $(o_i, g_i, a_i, r_i, o_{i+1})$ is appended to \mathcal{M}_t , progressively enriching the task-specific knowledge base.

Beyond storage, \mathcal{M}_t is continuously refined through reinforcement learning (RL), which estimates the long-term utility of experiences and guides policy improvement. Specifically, a Q-value function $Q(g, o, a)$ is maintained and updated using the Bellman target:

$$Q'(g, o_i, a_i) = r_i + \gamma \max_a Q(g, o_{i+1}, a), \tag{1}$$

where γ is the discount factor. For tasks with large or unbounded action spaces (e.g., natural language), the \max_a operator is approximated over a subset of observed actions. The Q-function is then updated using the standard Q-learning rule Watkins & Dayan (1992):

$$Q(g, o_i, a_i) \leftarrow (1 - \alpha)Q(g, o_i, a_i) + \alpha Q'(g, o_i, a_i), \tag{2}$$

where α is the learning rate.

The enhanced \mathcal{M}_t thus fulfills two complementary roles: (i) providing task-specific experiences that can be retrieved and encoded into LLM prompts for context-aware decision-making, and (ii) supporting policy refinement through Q-value updates, enabling the agent to adapt more effectively over time.

3.3 DYNAMIC MEMORY RETRIEVAL MECHANISM

While the source memory \mathcal{M}_s provides transferable priors and the target memory \mathcal{M}_t accumulates task-specific knowledge, their utility depends critically on how relevant experiences are retrieved.

To this end, we propose a *Dynamic Memory Retrieval Mechanism (DMRM)* that adaptively balances cross-task transfer and task-specific specialization.

At each decision step i , given the current observation–goal pair (o_i, g_i) , the agent ranks candidate entries $(o_{s/t}, g_{s/t}, a_{s/t}, r_{s/t})$ from $\mathcal{M}_s \cup \mathcal{M}_t$ using contextual similarity:

$$Sim = f((o_i, g_i), (o_{s/t}, g_{s/t})), \quad (3)$$

where $f(\cdot)$ is the cosine similarity between sentence embeddings. The top- k most relevant experiences are selected to guide inference.

To satisfy the input constraint of the underlying LLM, the number of retrieved entries k is dynamically adjusted. Let $\text{len}(\cdot)$ denote token length and C_{\max} the model’s context window. Retrieval proceeds until:

$$\sum_{j=1}^k \text{len}(m_j) + \text{len}(o_i, g_i) \leq C_{\max}, \quad (4)$$

ensuring that retrieval remains length-aware and maximally informative.

Beyond relevance and length, DMRM introduces an *adaptive source control* strategy to mitigate negative transfer. Specifically, once \mathcal{M}_t contains at least n high-quality successful trajectories, retrieval from \mathcal{M}_s is reduced or disabled:

$$|T_s| \geq n \Rightarrow \text{disable source retrieval},$$

where T_s is the set of successful trajectories in \mathcal{M}_t and n is a configurable threshold. This mechanism allows the agent to rely on transferable priors in the early stage of adaptation, while gradually shifting to task-specific experiences as evidence accumulates.

Finally, the retrieved experiences are serialized into natural language using a unified template (Figure 3), concatenated with the current (o_i, g_i) , and passed to the LLM for action generation. This retrieval process ensures that decision-making is both context-aware and adaptively balanced between cross-task knowledge transfer and in-task learning.

4 EXPERIMENT

In this section, we empirically evaluate the proposed method on the WebShop benchmark, comparing its performance with standard RL baselines and state-of-the-art LLM-based agents.

4.1 EXPERIMENTAL SETUPS

4.1.1 EVALUATION ENVIRONMENT

In this study, the WebShop Yao et al. (2022) serves as the experimental environment, offering a realistic simulation of multi-step online shopping via a web interface. It comprises over 1,000,000 product descriptions and 12,000 human-written instructions, supporting rigorous evaluation of instruction-following agents. In each task, the agent receives a user instruction and interacts with structured web pages via a discrete action set to identify a matching product. Only the five most recent interaction steps are visible, increasing the challenge of long-horizon reasoning. Episodes terminate upon product selection or after 15 steps, with a scalar reward in $[0, 1]$ reflecting task success.

WebShop spans diverse domains (e.g., beauty, electronics, grocery), with consistent UI structure but varying semantics and reasoning demands. We treat each domain as a separate task and assess cross-domain generalization by designating one as the source and others as targets. Following REMEMBERER Zhang et al. (2023), we adopt its setup: training on 10 randomly sampled tasks (disjoint from test tasks) for 3 epochs, repeated over five source–target pairs.

Memory retrieval employs a two-level similarity metric. Instruction-level similarity (f_g) is computed using SentenceTransformer all-MiniLM-L12-v2 Reimers & Gurevych (2019), while observation-level similarity (f_o) leverages the fixed layout structure of WebShop pages, assigning each to one of four predefined types via a lookup table. The final retrieval score is computed as in Equation (3).

4.1.2 COMPARED METHODS

To evaluate the effectiveness of our method, we compare it against several baselines. First, we consider an **LLM-only** setting, where a static GPT-3.5¹ or GPT-4² model is used with two fixed exemplars, without memory augmentation or task adaptation. We also include **ReAct** Yao et al. (2023), a strong baseline that interleaves reasoning and actions but relies solely on current observations and instructions, without leveraging past experiences. In addition, we compare with **REMEMBERER** Zhang et al. (2023), an LLM-based agent with memory-augmented RL that retrieves two fixed exemplars from its experience memory but lacks both cross-task knowledge transfer and adaptive memory selection during execution. Finally, we report results for standard **RL**, **IL**, and **IL+RL** agents, following Yao et al. (2023). To ensure a fair comparison, we align most hyperparameters with REMEMBERER Zhang et al. (2023). Unlike REMEMBERER, our method employs (1) a source memory storing transferable experiences and (2) a dynamic retrieval mechanism that selects relevant cases based on the current state and task progression, enabling more adaptive and context-aware memory utilization.

4.1.3 EVALUATION METRICS

We evaluate the effectiveness of our memory-augmented LLM agent with cross-task experience learning on the WebShop benchmark using two standard metrics. The **average score** is a scalar feedback in the range $[0, 1]$, where higher values indicate better alignment between the selected product and the user instruction. The **success rate**, ranging from $[0, 1]$, indicates the all task completion, with each task counted as a success (1) or failure (0). In both metrics, higher values indicate better agent performance. These metrics are widely adopted in prior works on instruction-following agents Yao et al. (2023); Zhang et al. (2023) because they effectively capture both qualitative and quantitative aspects of agent performance, balancing reward precision and task completion robustness.

4.2 RESULTS AND DISCUSSIONS

All agents in our experiments are powered by an instruction-tuned LLM that remains frozen throughout evaluation, without any gradient-based fine-tuning. Instead, their decision-making relies on carefully designed prompt engineering to elicit reasoning and adaptation behaviors.

Table 1: Overall performance comparison across baseline and proposed methods on the WebShop benchmark. Metrics include average instruction satisfaction score and task success rate.

Method	Avg Score	Success Rate
LLM only	0.55	0.29
ReAct	0.66	0.36
RMMBR. (w/o transfer)	0.68	0.38
Ours (Mem+Transfer)	0.70	0.40
RL	0.55	0.18
IL	0.60	0.29
IL+RL	0.62	0.29

Table 1 reports performance on the target category in terms of average score and success rate. Our method with cross-task transfer consistently outperforms all baselines, including LLM-only prompting, ReAct Yao et al. (2023), REMEMBERER without transfer Zhang et al. (2023), and standard RL, IL, and IL+RL approaches. It achieves the highest average score (0.70) and success rate (0.40), underscoring the benefits of incorporating transferable experience into LLM-based agents. These gains can be attributed to two key components of our design. First, the source experience memory provides access to transferable knowledge from related tasks, reducing the reliance on large amounts of task-specific interaction data and thereby improving data efficiency. Second, the dynamic memory retrieval mechanism adaptively balances source and target experiences, allowing the agent to

¹<https://platform.openai.com/docs/models/gpt-3.5-turbo>

²<https://platform.openai.com/docs/models/gpt-4-turbo>

exploit transferable knowledge when beneficial while specializing on target-specific details as interaction progresses. Together, these mechanisms enable the agent to generalize more effectively across domains, yielding robust performance even under limited prior task experience.

Table 2: Transfer performance under **different target tasks**. For each method, the first row reports the mean scores, and the second row reports the success rates. Task abbreviations: **fa** = fashion, **be** = beauty, **el** = electronics, **gr** = grocery.

Method	$\mathcal{M}_{t_{be}}$	$\mathcal{M}_{t_{el}}$	$\mathcal{M}_{t_{gr}}$	Avg
LLM only	0.55	0.53	0.57	0.54
LLM only	0.28	0.26	0.30	0.28
RMMBR. (w/o transfer)	0.68	0.66	0.70	0.68
RMMBR. (w/o transfer)	0.38	0.37	0.39	0.38
	$\mathcal{M}_{s_{fa}} + \mathcal{M}_{t_{be}}$	$\mathcal{M}_{s_{fa}} + \mathcal{M}_{t_{el}}$	$\mathcal{M}_{s_{fa}} + \mathcal{M}_{t_{gr}}$	Avg
Ours (transfer)	0.71	0.70	0.69	0.70
Ours (transfer)	0.41	0.39	0.40	0.40

Table 3: Transfer performance under **different source tasks**. For each method, the first row reports the mean scores, and the second row reports the success rates. Task abbreviations: **fa** = fashion, **be** = beauty, **el** = electronics, **gr** = grocery.

Method	$\mathcal{M}_{t_{be}}$	$\mathcal{M}_{t_{be}}$	$\mathcal{M}_{t_{be}}$	Avg
LLM only	0.55	-	-	0.55
LLM only	0.28	-	-	0.28
RMMBR.(w/o transfer)	0.68	-	-	0.68
RMMBR.(w/o transfer)	0.38	-	-	0.38
	$\mathcal{M}_{s_{fa}} + \mathcal{M}_{t_{be}}$	$\mathcal{M}_{s_{el}} + \mathcal{M}_{t_{be}}$	$\mathcal{M}_{s_{gr}} + \mathcal{M}_{t_{be}}$	Avg
Ours (transfer)	0.70	0.72	0.71	0.71
Ours (transfer)	0.40	0.41	0.40	0.40

Tables 2 and 3 present cross-task transfer results across four product domains: *fashion* (fa), *beauty* (be), *electronics* (el), and *grocery* (gr). We examine two complementary transfer settings. Table 2 varies the *target task* while fixing the source memory, evaluating the agent’s ability to adapt to unseen domains using knowledge from a single source. Table 3 varies the *source task* while keeping the target fixed, testing robustness when transferable knowledge is drawn from heterogeneous sources. In both cases, our method surpasses all baselines, achieving consistently higher scores and success rates. The observed improvements stem from the interaction of two key components. First, the *source experience memory* enables the agent to reuse transferable strategies across domains, reducing the need for extensive task-specific exploration. This explains why adaptation remains effective even when source and target domains are semantically distant. Second, the *dynamic retrieval mechanism* regulates how source and target memories are weighted during inference, allowing the agent to selectively transfer useful patterns while avoiding negative transfer. This adaptive balance underlies the strong generalization observed in both transfer directions, demonstrating that our approach can efficiently exploit prior knowledge while preserving target-task specialization.

4.3 ABLATION STUDY

4.3.1 EFFECT OF SUPPRESSION THRESHOLD n .

As detailed in Section *Dynamic Memory Retrieval Mechanism*, the parameter n sets a minimum coverage threshold that determines when to suppress retrieval from source memory \mathcal{M}_s , thereby shifting focus to target memory \mathcal{M}_t once sufficient task-specific experience is available. This design improves retrieval efficiency and mitigates negative transfer. We evaluate $n \in 1, 2, 3, 4$ to capture different adaptation stages: smaller n accelerates reliance on \mathcal{M}_t , while larger n prolongs dependence on \mathcal{M}_s . As shown in Table 4, $n = 2$ yields the best performance (0.70 average score,

Table 4: Performance under different memory transition thresholds n .

Setting	Avg Score	Success Rate
$n = 1$	0.67	0.38
$n = 2$	0.70	0.40
$n = 3$	0.71	0.39
$n = 4$	0.71	0.38

0.40 success rate). A smaller value ($n = 1$) detaches from \mathcal{M}_s too early, reducing performance (0.67 / 0.38), while larger values ($n = 3, 4$) delay suppression and introduce noise, leading to marginal or degraded results. Overall, a moderate threshold ($n = 2$) achieves the best balance between transfer and adaptation.

4.3.2 EFFECT OF TRANSFER AND DYNAMIC MEMORY RETRIEVAL.

Table 5: Performance comparison across ablation settings.

Setting	Avg Score	Success Rate
RMMBR. (w/o transfer)	0.68	0.38
Ours (w/o DMRM)	0.69	0.38
Ours (Full DMRM)	0.70	0.40

To further isolate the contributions of transfer learning and the proposed retrieval mechanism, we compare three settings: (i) **No Transfer**, where the agent relies solely on target memory \mathcal{M}_t without access to source experiences; (ii) **w/o DMRM**, where both source and target memories are available but retrieval follows a static strategy without adaptive suppression; and (iii) **Full DMRM**, our proposed approach with dynamic retrieval. As shown in Table 5, No Transfer yields the weakest performance (average score: 0.68, success rate: 0.38), indicating that access to prior experiences is crucial for enhancing target-task learning. Adding source memory without adaptive regulation (w/o DMRM) provides only marginal improvement (0.69 / 0.38), as indiscriminate retrieval introduces noise and negative transfer that offset potential gains. In contrast, Full DMRM achieves the best performance (0.70 / 0.40). This improvement arises from the *dynamic retrieval mechanism*, which adaptively balances the use of source and target experiences, amplifying beneficial transfer while suppressing irrelevant knowledge. These results confirm that both components are indispensable: cross-task memory supplies transferable knowledge, while dynamic retrieval ensures its effective and selective integration.

5 CONCLUSION

In this work, we have developed a memory-augmented LLM-based agent with cross-task experience learning to enhance adaptability and generalization in multi-turn decision-making. Our design has extended conventional task-specific memory with a source experience memory that stores transferable knowledge from related but distinct tasks, and has introduced a Dynamic Memory Retrieval Mechanism that adaptively balances task-specific and cross-task experiences throughout interaction. We have extensively evaluated our approach on the WebShop benchmark, where it has consistently outperformed strong memory-augmented baselines in both task success rate and generalization. These findings have demonstrated that integrating transferable memory with adaptive retrieval not only has improved data efficiency but also has enabled more robust and flexible decision-making in complex, multi-turn environments.

REFERENCES

Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Daniel Ho, Jasmine Hsu, Julian Ibarz, Brian Ichter, Alex Irpan, Eric Jang, Rosario Jauregui Ruano, Kyle Jeffrey, Sally

- 486 Jesmonth, Nikhil J Joshi, Ryan Julian, Dmitry Kalashnikov, Yuheng Kuang, Kuang-Huei Lee,
487 Sergey Levine, Yao Lu, Linda Luu, Carolina Parada, Peter Pastor, Jornell Quiambao, Kanishka
488 Rao, Jarek Rettinghouse, Diego Reyes, Pierre Sermanet, Nicolas Sievers, Clayton Tan, Alexander
489 Toshev, Vincent Vanhoucke, Fei Xia, Ted Xiao, Peng Xu, Sichun Xu, Mengyuan Yan, and Andy
490 Zeng. Do as i can, not as i say: Grounding language in robotic affordances, 2022.
- 491 Ekin Akyürek, Dale Schuurmans, Jacob Andreas, Tengyu Ma, and Denny Zhou. What learning
492 algorithm is in-context learning? investigations with linear models, 2023.
- 493
- 494 Ning Bian, Xianpei Han, Le Sun, Hongyu Lin, Yaojie Lu, Ben He, Shanshan Jiang, and Bin Dong.
495 Chatgpt is a knowledgeable but inexperienced solver: An investigation of commonsense problem
496 in large language models. In *Proceedings of the 2024 Joint International Conference on Computa-*
497 *tional Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 3098–3110,
498 Torino, Italia, 2024. ELRA and ICCL.
- 499 Xinshi Chen, Shuang Li, Hui Li, Shaohua Jiang, Yuan Qi, and Le Song. Generative adversarial user
500 model for reinforcement learning based recommendation system. In *International conference on*
501 *machine learning*, pp. 1052–1061, Long Beach, CA, USA, 2019. PMLR.
- 502
- 503 Yuqing Du, Olivia Watkins, Zihan Wang, Cédric Colas, Trevor Darrell, Pieter Abbeel, Abhishek
504 Gupta, and Jacob Andreas. Guiding pretraining in reinforcement learning with large language
505 models. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 8657–
506 8677, Honolulu, Hawaii, USA, 2023. PMLR.
- 507 Kevin A. Fischer. Reflective linguistic programming (rlp): A stepping stone in socially-aware agi
508 (socialagi), 2023.
- 509
- 510 Chenxu Hu, Jie Fu, Chenzhuang Du, Simian Luo, Junbo Zhao, and Hang Zhao. Chatdb: Augment-
511 ing llms with databases as their symbolic memory, 2023.
- 512 Jie Huang and Kevin Chen-Chuan Chang. Towards reasoning in large language models: A survey.
513 In *Findings of the Association for Computational Linguistics, ACL 2023*, pp. 1049–1065, Toronto,
514 Canada, 2023. Association for Computational Linguistics (ACL).
- 515
- 516 Wenlong Huang, Fei Xia, Ted Xiao, Harris Chan, Jacky Liang, Pete Florence, Andy Zeng, Jonathan
517 Tompson, Igor Mordatch, Yevgen Chebotar, Pierre Sermanet, Tomas Jackson, Noah Brown, Linda
518 Luu, Sergey Levine, Karol Hausman, and brian ichter. Inner monologue: Embodied reasoning
519 through planning with language models. In *Proceedings of The 6th Conference on Robot Learn-*
520 *ing*, pp. 1769–1782, Auckland, New Zealand, 2023a. PMLR.
- 521
- 522 Ziheng Huang, Sebastian Gutierrez, Hemanth Kamana, and Stephen Macneil. Memory sandbox:
523 Transparent and interactive memory management for conversational agents. In *Adjunct Proceed-*
524 *ings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 97,
New York, NY, USA, 2023b. Association for Computing Machinery.
- 525
- 526 Ehud Karpas, Omri Abend, Yonatan Belinkov, Barak Lenz, Opher Lieber, Nir Ratner, Yoav Shoham,
527 Hofit Bata, Yoav Levine, Kevin Leyton-Brown, Dor Muhlgay, Noam Rozen, Erez Schwartz,
528 Gal Shachaf, Shai Shalev-Shwartz, Amnon Shashua, and Moshe Tenenholz. Mrkl systems: A
529 modular, neuro-symbolic architecture that combines large language models, external knowledge
sources and discrete reasoning, 2022.
- 530
- 531 Ali Modarressi, Ayyoob Imani, Mohsen Fayyaz, and Hinrich Schütze. Ret-llm: Towards a general
read-write memory for large language models, 2024.
- 532
- 533 Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christopher
534 Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, Xu Jiang, Karl Cobbe, Tyna Eloundou,
535 Gretchen Krueger, Kevin Button, Matthew Knight, Benjamin Chess, and John Schulman. Webgpt:
536 Browser-assisted question-answering with human feedback, 2022.
- 537
- 538 Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and
539 Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Pro-*
ceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pp.
2:1–2:22, San Francisco, CA, USA, 2023. Association for Computing Machinery.

- 540 Shreyas Sundara Raman, Vanya Cohen, Eric Rosen, Ifrah Idrees, David Paulius, and Stefanie Tellex.
541 Planning with large language models via corrective re-prompting. In *NeurIPS 2022 Foundation*
542 *Models for Decision Making Workshop*, New Orleans, LA, USA, 2022. NeurIPS Foundation.
- 543
544 Krishan Rana, Jesse Haviland, Sourav Garg, Jad Abou-Chakra, Ian Reid, and Niko Suenderhauf.
545 Sayplan: Grounding large language models using 3d scene graphs for scalable robot task plan-
546 ning. In *Proceedings of The 7th Conference on Robot Learning*, pp. 23–72, Atlanta, USA, 2023.
547 PMLR.
- 548 Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-
549 networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language*
550 *Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-*
551 *IJCNLP)*, pp. 3982–3992, Hong Kong, China, 2019. Association for Computational Linguistics.
- 552
553 Jingqing Ruan, Yihong Chen, Bin Zhang, Zhiwei Xu, Tianpeng Bao, Guoqing Du, Shiwei Shi,
554 Hangyu Mao, Ziyue Li, Xingyu Zeng, and Rui Zhao. Tptu: Large language model-based ai
555 agents for task planning and tool usage, 2023.
- 556 Timo Schick, Jane Dwivedi-Yu, Roberto Dessi, Roberta Raileanu, Maria Lomeli, Eric Hambro,
557 Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can
558 teach themselves to use tools. In *Advances in Neural Information Processing Systems*, pp. 68539–
559 68551, New Orleans, Louisiana, USA, 2023. Curran Associates, Inc.
- 560
561 Melanie Sclar, Sachin Kumar, Peter West, Alane Suhr, Yejin Choi, and Yulia Tsvetkov. Minding
562 language models’ (lack of) theory of mind: A plug-and-play multi-character belief tracker, 2023.
- 563
564 Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueting Zhuang. Hug-
565 ginggpt: Solving ai tasks with chatgpt and its friends in hugging face. In *Advances in Neural*
566 *Information Processing Systems*, pp. 38154–38180, New Orleans, Louisiana, USA, 2023. Curran
567 Associates, Inc.
- 568 Zhenmei Shi, Junyi Wei, Zhuoyan Xu, and Yingyu Liang. Why larger language models do in-context
569 learning differently? In *Proceedings of the 41st International Conference on Machine Learning*,
570 pp. 1831:1–1831:23, Vienna, Austria, 2024. JMLR.org.
- 571
572 Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflex-
573 ion: language agents with verbal reinforcement learning. In *Advances in Neural Information*
574 *Processing Systems*, pp. 8634–8652, New Orleans, LA, USA, 2023. Curran Associates, Inc.
- 575
576 Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter
577 Fox, Jesse Thomason, and Animesh Garg. Progprompt: program generation for situated robot
578 task planning using large language models. *Autonomous Robots*, 47:999–1012, 2023.
- 579
580 Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel M. Ziegler, Ryan Lowe, Chelsea Voss, Alec
581 Radford, Dario Amodei, and Paul Christiano. Learning to summarize with human feedback. In
582 *Advances in Neural Information Processing Systems*, pp. 3008–3021, Vancouver, Canada, 2020.
Curran Associates, Inc.
- 583
584 Richard S Sutton, Andrew G Barto, et al. (eds.). *Reinforcement learning: An introduction*. MIT
585 press Cambridge, Cambridge, MA, 1998.
- 586
587 Qingyue Wang, Yanhe Fu, Yanan Cao, Shuai Wang, Zhiliang Tian, and Liang Ding. Recursively
588 summarizing enables long-term dialogue memory in large language models. *Neurocomputing*,
639:130193, 2025.
- 589
590 Zihao Wang, Shaofei Cai, Guanzhou Chen, Anji Liu, Xiaojian Ma, Yitao Liang, and Team Craft-
591 Jarvis. Describe, explain, plan and select: interactive planning with large language models enables
592 open-world multi-task agents. In *Proceedings of the 37th International Conference on Neural In-*
593 *formation Processing Systems*, pp. 1480, Red Hook, NY, USA, 2023. Curran Associates Inc.
- Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8:279–292, 1992.

- 594 Shunyu Yao, Howard Chen, John Yang, and Karthik Narasimhan. Webshop: Towards scalable
595 real-world web interaction with grounded language agents. In *Advances in Neural Information*
596 *Processing Systems*, pp. 20744–20757, New Orleans, LA, USA, 2022. Curran Associates, Inc.
597
- 598 Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao.
599 React: Synergizing reasoning and acting in language models. In *Proceedings of the 11th In-*
600 *ternational Conference on Learning Representations (ICLR)*, pp. 1–18, Kigali, Rwanda, 2023.
601 Conference on Learning Representations.
- 602 Danyang Zhang, Lu Chen, Situo Zhang, Hongshen Xu, Zihan Zhao, and Kai Yu. Large language
603 models are semi-parametric reinforcement learning agents. In *Advances in Neural Information*
604 *Processing Systems*, pp. 78227–78239, New Orleans, LA, USA, 2023. Curran Associates, Inc.
605
- 606 Andrew Zhu, Lara Martin, Andrew Head, and Chris Callison-Burch. Calypso: Llms as dungeon
607 master’s assistants. *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive*
608 *Digital Entertainment*, 19:380–390, 2023a.
- 609 Xizhou Zhu, Yuntao Chen, Hao Tian, Chenxin Tao, Weijie Su, Chenyu Yang, Gao Huang, Bin Li,
610 Lewei Lu, Xiaogang Wang, Yu Qiao, Zhaoxiang Zhang, and Jifeng Dai. Ghost in the minecraft:
611 Generally capable agents for open-world environments via large language models with text-based
612 knowledge and memory, 2023b.

613
614
615

A APPENDIX

616
617

B ETHICS STATEMENT

618
619
620
621
622
623
624
625

This work complies with the ICLR Code of Ethics. Our study does not involve human participants or animal experimentation. All datasets employed are publicly available and used strictly in accordance with their licensing and usage guidelines, thereby avoiding any violation of privacy or intellectual property. We have carefully considered potential risks of bias or unintended harm. No personally identifiable information was used, and no experiments were conducted that could raise privacy, fairness, or security concerns. We are committed to conducting and presenting our research with transparency, integrity, and respect for ethical standards.

626
627

C LLM USAGE

628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647

Large Language Models (LLMs) were used solely to assist in the preparation of this manuscript. Their contributions were limited to language refinement tasks such as improving readability, enhancing clarity, and ensuring stylistic consistency with academic writing standards. The LLM was not involved in conceptual development, experimental design, implementation, or analysis. All scientific ideas, methodologies, and results were conceived and validated independently by the authors. The authors take full responsibility for the final content of this paper. We have carefully reviewed all text produced with LLM assistance to ensure accuracy, originality, and compliance with ethical guidelines, thereby avoiding plagiarism or scientific misconduct. The LLM was employed strictly as a writing aid, without influencing the research substance.