From Facts to Insights: A Persona-Driven Dual Memory Framework and Dataset for Role-Playing Agents

Rongsheng Zhang[1,2*], Ruofan Hu[1,2*], Weijie Chen[2], Jiji Tang[2], Junnan Ren[2], Wanying Wu[1], Xunuoyan Chen[1], Tangjie Lv[2], Tao Jin[1], Zhou Zhao[1†]

[1]Zhejiang University [2]Fuxi AI Lab, Netease Inc.

njuzrs@163.com ruofanhu@zju.edu.cn

Abstract

While role-playing agents excel in short-term interactions, long-term conversations overwhelm context windows, motivating external memory frameworks. Current systems typically rely on persona-agnostic summarization, which records facts without persona-specific interpretation, yielding generic responses that compromise persona fidelity. To bridge this gap, we introduce RoleMemo, a dataset featuring four reasoning tasks where the factual fragments must be interpreted through the persona to reach the correct answer. Evaluation on RoleMemo exposes critical limitations of persona-agnostic frameworks. We thus propose DUALMEM, which decouples memory into two streams: factual cognition and persona-conditioned insight. Trained through Supervised Fine-Tuning (SFT) and Reinforcement Learning (RL), our framework with a **4B-parameter** model outperforms zero-shot persona-agnostic frameworks powered by DeepSeek-V3.2 for sustained persona fidelity. Our resources are available at https://github.com/Collab-Gen/RoleMemo.

1 Introduction

The rapid advancement of large language models (LLMs) has propelled role-playing agents beyond superficial style transfer toward deep emotional companionship (Zhang et al., 2025b; Wang et al., 2025b; Zhang et al., 2025a). To achieve this, agents must filter the interaction history through their assigned persona, retaining what matters for in-character response. However, as conversation grows, the underlying LLM's attention is diluted by trivial information, undermining this persona filter and compromising fidelity (Wu et al., 2025a). Existing methods tackle this challenge by adopting an agentic memory framework, where a persona-agnostic model summarizes past conversations into neutral facts, for the role-playing agent to retrieve and utilize at inference time.

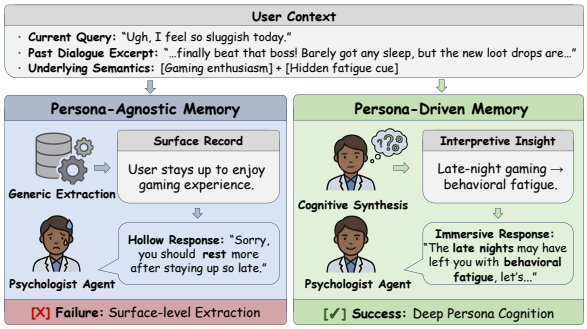

Figure 1: Persona-agnostic memory stores the neutral fact, forcing the agent to reinterpret it at inference time and fall back to a generic reply. Persona-driven memory stores a persona-conditioned interpretation grounded in the fact, enabling an in-character response.

However, role-playing agents are designed to emulate human cognition, where memory is an active cognitive process rather than a neutral fact repository. Cognitive science describes this as reconstructive memory, where past experience is interpreted through one's perspective rather than literally recalled (Bartlett, 1995). As Fig. 1 illustrates, current persona-agnostic memory records "late-night gaming" as a neutral fact, while a psychologist agent's memory should instead filter and store it as "behavioral fatigue". Without that persona-conditioned interpretation in memory, the agent retrieves only the surface fact at inference time and must reinterpret it from scratch (He et al., 2025). The reconstruction is unreliable (He et al., 2025; Chen et al., 2026), as the persona cues are diluted in the retrieved long context. The psychologist then falls back to a generic "you should rest more" reply that misses the persona-conditioned reading.

Current evaluation systems for agentic memory further mask this flaw. Mainstream benchmarks primarily evaluate fact retrieval through needle-in-a-haystack paradigms (Maharana et al., 2024; Hu et al., 2025b). Persona-agnostic frameworks eas-

[*]Equal contribution.
[†]Corresponding author.

Findings of the Association for Computational Linguistics: EMNLP 2026, pages 5666–5686
October 24-29, 2026 ©2026 Association for Computational Linguistics

Table 1: Comparison of RoleMemo with existing memory benchmarks. RoleMemo features: (1) Cross-Session Reasoning linking clues across distant turns, (2) Deep Persona Interpretation mapping facts to persona-specific insights, (3) In-Character Evaluation of role-playing fidelity, and (4) LLM Fine-Tuning data for memory construction. The **Context Length** column for RoleMemo reports the maximum evaluation length; training uses 32k contexts.

Dataset	Persona Count	Topic Count	Context Length	Total Query	Cross-Session Reasoning	Deep Persona Interpretation	In-Character Evaluation	LLM Fine-Tuning Data
LoCoMo (Maharana et al., 2024)	100	N/A	9k	7,512	×	×	×	×
LongMemEval (Wu et al., 2025a)	N/A	N/A	**1.5M**	500	✓	×	×	×
MemAgentBench (Hu et al., 2025b)	N/A	N/A	1.44M	2,071	✓	×	×	×
PrefEval (Zhao et al., 2025)	N/A	20	100k	3,000	×	×	×	×
PersonaMem-v1 (Jiang et al., 2025a)	180	15	1.0M	2,700	✓	×	×	×
PersonaMem-v2 (Jiang et al., 2025b)	1,000	335	128k	5,000	✓	×	×	✓
RoleMemo (Ours)	**2,052**	**1,702**	256k	**20,244**	✓	✓	✓	✓

ily succeed by extracting universally salient facts, such as phone numbers or specific dates. As a result, these facts are sufficient to resolve queries regarding information localization tasks, bypassing the need for persona-conditioned interpretation.

To bridge this gap, we introduce RoleMemo, a large-scale dataset for evaluating agentic memory in role-playing. RoleMemo targets what prior benchmarks miss: persona-conditioned reconstructive memory over long, cross-session histories, which we denote *insight*. As Tab. 1 illustrates, RoleMemo spans 1,702 topics with conversation histories up to 256k tokens, comprising a training set of 1,914 personas with 18,088 queries and an evaluation set of 138 personas with 2,156 queries. It features four query types inspired by social cognition frameworks (Epley et al., 2004). To evaluate these tasks, we measure retrieval accuracy and memory-driven response quality with in-character evaluation following CharacterEval (Tu et al., 2024).

Evaluations on RoleMemo reveal that persona-agnostic memory frameworks struggle with these specialized tasks. We thus propose DualMem, a persona-driven dual memory framework for high-fidelity role-playing. Specifically, DualMem first captures objective events as *factual cognition*, then derives persona-conditioned interpretations from these facts as *insight cognition*. A memory-specialized 4B model trained via SFT (Ouyang et al., 2022) and RL (Guo et al., 2025) further drives this construction, yielding DualMem-SFT and DualMem-RL that outperform zero-shot persona-agnostic frameworks on role-playing quality.

Our contributions are summarized as follows:

- We construct RoleMemo, the first benchmark for persona-conditioned reconstructive memory in role-playing scenarios.
- We evaluate nine zero-shot persona-agnostic frameworks on RoleMemo, revealing a structural insight bottleneck that persists regardless of driving model scale.
- We propose DualMem, which decouples memory into factual and insight cognition. With training, our 4B memory model outperforms baselines driven by 685B models on role-playing quality.

2 Related Work

2.1 Persona-Agnostic Memory Frameworks

Role-playing agents aim to provide emotional companionship (Zhang et al., 2025b; Wang et al., 2025b; Zhang et al., 2025a). However, as interaction history expands, these models suffer from attention dilution and struggle to maintain persona fidelity over long contexts. Early solutions employ Retrieval-Augmented Generation (RAG) to search uncompressed conversation histories (Zerhoudi and Granitzer, 2024; Yang et al., 2025b; Wang et al., 2025c; Park et al., 2025), imposing storage overhead. Recent studies shift toward agentic memory frameworks (An et al., 2026). They use persona-agnostic models to iteratively extract facts from dialogues and update prior memory through hierarchical structuring (Zhang et al., 2026; Wang et al., 2025a; Wang and Chen, 2025), lightweight compression (Fang et al., 2025; Liu et al., 2026), and task-driven predictive modeling (Chhikara et al., 2025; Kim et al., 2025; Zheng et al., 2025). Nevertheless, these approaches prioritize salient facts over persona-specific nuances, failing to provide the interpretation for high-fidelity role-playing.

A parallel line of work targets long-horizon task-solving agents, where higher-level memory is constructed by reflecting on execution traces like Reflexion (Shinn et al., 2023; Zhai et al., 2025; Wu et al., 2025b). Such memory abstracts experience for future task completion, a goal distinct from persona-conditioned interpretation in role-playing memory. We compare them in detail in App. A.

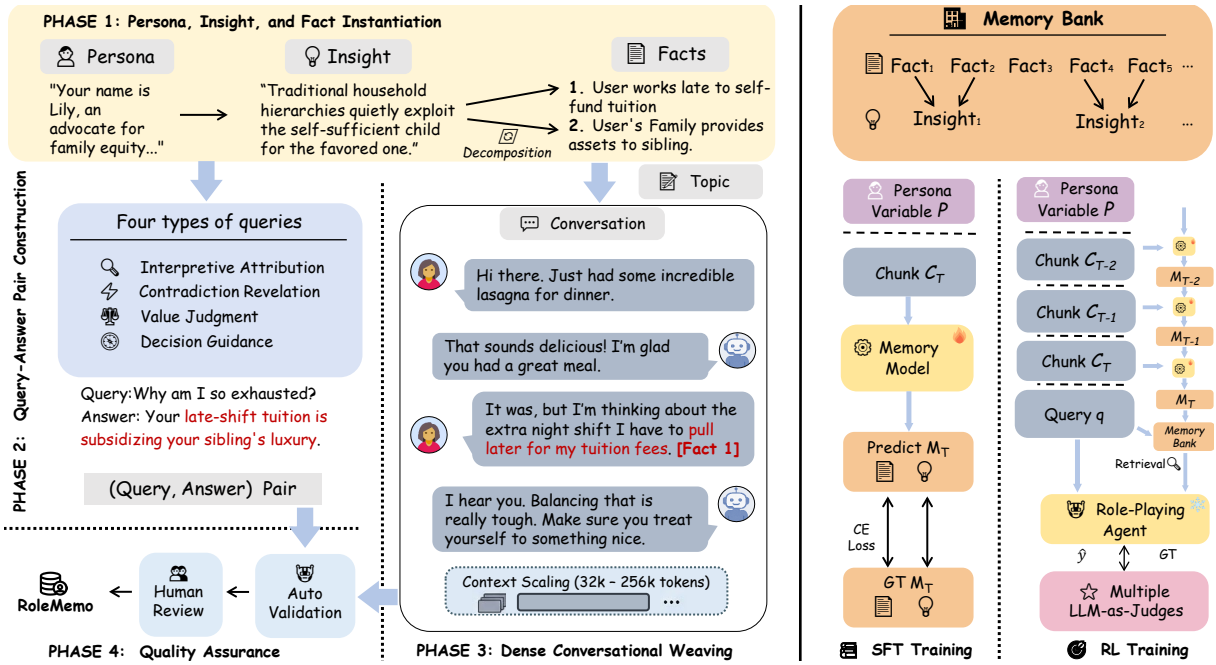

Figure 2: The RoleMemo dataset and DUALMEM framework. Left: The four-phase dataset construction pipeline. We generate queries that require agents to synthesize scattered facts via persona-driven insights to formulate answers. Right: Our proposed DUALMEM framework. It maintains a unified Memory Bank that decouples factual and insight cognition while preserving their links. We explore two training mechanisms on RoleMemo: SFT and RL.

2.2 Memory and Role-Playing Benchmarks

Benchmarking persona-driven agentic memory also remains underexplored. Current memory benchmarks (Maharana et al., 2024; Wu et al., 2025a; Pakhomov et al., 2025; Hu et al., 2025b) evaluate agentic memory frameworks primarily through explicit fact retrieval. By adopting a needle-in-a-haystack paradigm (Nelson et al., 2024), they prioritize literal matching between queries and keys, failing to evaluate how memory informs persona-conditioned responses. Conversely, role-playing benchmarks (Tu et al., 2024; Chen et al., 2024; Zhou et al., 2025) emphasize persona fidelity but confine memory evaluation to narrow conversation windows. They measure short-term conversational fluency rather than the retention of long contexts. This gap necessitates a benchmark to evaluate the fidelity of personas across extended interactions.

3 RoleMemo Dataset Construction

We introduce RoleMemo, a dataset in which queries require reconstructing scattered facts into insights through persona-driven interpretation. As illustrated in the left part of Fig. 2, the construction process consists of four phases. We employ DeepSeek-V3.2 (Liu et al., 2025) for dataset generation, detailing prompts and parameters in App. B.

Effective evaluation of persona-driven interpretation requires insights that capture persona-specific stances grounded in lived experience. We instantiate 2,052 personas across 23 domains, each defined by professional backgrounds and hobbies, with details in App. C. Since identity-level profiles do not expose persona-specific judgments, we instantiate each persona as a set of concrete insights that can be queried and evaluated. For instance, an advocate for family equity might hold that "self-sufficient children are exploited for favored siblings," a judgment that only this persona's background would produce. This process yields 26,636 insights.

To ground these insights, we generate discrete factual fragments that remain neutral in isolation but reveal meaning through persona-driven interpretation. For the insight regarding family equity, the factual fragments include "works late to self-fund tuition" and "the family provides assets to a sibling". These pairings of facts and insights form the core unit of RoleMemo. Answering any query requires the agent to interpret retrieved facts rather than relying on surface-level retrieval.

3.1 Query-Answer Pair Construction

Using these fact-insight pairs, we design queries with ground-truth responses across four types inspired by social cognition frameworks (Epley et al., 2004). Each query requires insight instead of scat-

tered facts to generate answers. Specifically, *Interpretive Attribution* infers unstated motives from behavioral patterns, *Contradiction Revelation* detects inconsistencies between values and actions, *Value Judgment* evaluates situations through persona ethics, and *Decision Guidance* recommends actions aligned with principles.

As shown in Fig. 2, the query "Why am I so exhausted?" belongs to *Interpretive Attribution* type. No single fact reveals the underlying pattern, but combined they expose a motive recognizable only to a family equity advocate, producing the response "Your overtime is subsidizing the luxury of a sibling."

3.2 Dense Conversational Weaving

Facts must remain dormant within natural conversation until a query forces their retrieval. We embed these factual fragments into contexts by generating conversation histories through a layered process. The process begins by selecting topics from 1,702 topics aligned with the background of each persona, followed by the generation of conversational blocks comprising 24 turns between the persona and an interlocutor. To increase retrieval difficulty, we introduce multiple-entity distractions related to the selected topic within each block. Factual fragments are then injected under naturalistic constraints, ensuring they blend contextually, depend syntactically on surrounding utterances, and remain incidental to the focus. Finally, by interleaving evidence-rich blocks with casual conversation at varying densities, we scale histories from 32k to 256k tokens, with multiple queries per persona.

3.3 Quality Assurance

RoleMemo validates each instance through four automated checks. First, we verify **insight specificity** by confirming that insight requires the persona rather than generic domain knowledge. Second, we verify **memory-necessity** by prompting the role-playing agent to answer queries without the history, ensuring all queries require retrieval and interpretation rather than common-sense reasoning. Third, we enforce **difficulty control** by confirming the evidence appears only in single turns, preventing the exploitation of high-frequency patterns. Fourth, we guarantee **safety** by filtering all conversations for violence and privacy violations.

To validate the reliability of these automated checks, we randomly sampled 500 instances for human review across all criteria, achieving agreement with Cohen's $\kappa = 0.83$. By applying these checks to the entire dataset, we retain 20,244 queries along with the corresponding conversation histories.

4 Persona-Driven Memory Framework

4.1 Decoupled Dual Memory Framework

High-fidelity role-playing requires agents to extract and interpret conversation details through their persona. However, mainstream memory frameworks adopt a neutral perspective, reducing interactions to isolated facts without persona-conditioned interpretation. To address this gap, as shown in Fig. 2, we design DUALMEM around two complementary cognition types.

Specifically, *factual cognition* captures objective events and semantic details from conversation history, preserving the informational foundation maintained by existing approaches. *Insight cognition*, our core contribution, derives persona-driven interpretations from factual cognition. Each insight references its grounding facts to maintain interpretive traceability at inference time. These cognitions are jointly stored in a unified memory bank.

4.2 Persona-guided Memory Construction

We propose an incremental memory construction process to operationalize the dual framework. Following Yu et al. (2026), we partition the conversation history into fixed-size chunks and process them sequentially. Along this pipeline, we maintain a global memory bank $M = \bigcup_{i=1}^{T} M_i$, accumulating cognitions M_i generated at each step i.

To generate this step-wise cognition M_i, memory construction model f_θ processes persona profile p, chunk C_i, and prior memory $M_{<i}$ as prompt input. First, the model extracts factual cognition F_i, capturing objective events from C_i to ensure details relevant to the persona are retained. Then, the model constructs insight cognition I_i by interpreting F_i and relevant history from $M_{<i}$ through the persona. After processing all T chunks, a separate role-playing agent f_ϕ generates the response $\hat{y}$ to query q using cognition retrieved from the complete memory M. Formally:

$$\begin{aligned} M_i &= \{F_i, I_i\} = f_\theta(C_i, M_{<i}, p), \\ \hat{y} &= f_\phi(p, q, \mathcal{R}(M, q)), \end{aligned} \tag{1}$$

where p denotes the persona, F_i and I_i are factual and insight cognition at step i, and $\mathcal{R}(\cdot, \cdot)$ retrieves relevant cognition for query q.

Notably, during retrieval, when an insight cognition is retrieved, its linked factual cognitions are

also included to provide grounding evidence for f_ϕ. This design ensures interpretations remain anchored in factual evidence.

4.3 Training Mechanisms on RoleMemo

We explore SFT and RL to endow f_θ with memory construction capabilities from RoleMemo, yielding DUALMEM-SFT and DUALMEM-RL respectively, both initialized from the base model.

Under SFT, we train f_θ for dual memory operations through instruction tuning. Using conversation chunks from RoleMemo, we optimize two objectives. First, the model extracts factual cognition F_i from C_i and p. Second, given these facts and relevant information from $M_{<i}$, the model generates insight cognition I_i through persona p. Furthermore, to mitigate forced interpretation hallucinations, we incorporate manually constructed negative samples into the training data: when the evidence from $M_{<i}$ and F_i is insufficient, the target insight is set to null ($I_i = \varnothing$). Formalizing this decoupled generation, we maximize the log-likelihood across T independent chunks:

$$\mathcal{L}_{\text{SFT}}(\theta) = \sum_{i=1}^{T} \Big(\log P_\theta(F_i \mid C_i, p) + \log P_\theta(I_i \mid F_i, M_{<i}, p) \Big) \tag{2}$$

Under RL, we extend Yu et al. (2026)'s multi-turn conversation framework for dual memory construction. Given the entire long histories from RoleMemo, f_θ autonomously executes multiple rounds of cognition construction without intermediate supervision, generating a complete memory trajectory $\tau = \{M_i\}_{i=1}^{T}$, where $M_i = \{F_i, I_i\}$. The fixed role-playing agent f_ϕ then generates response $\hat{y}$ to query q using cognition retrieved from the memory bank M. GPT-5.1 evaluates responses using two reward functions: Format Compliance and In-Character Quality (Tu et al., 2024), yielding a combined scalar reward $R(\hat{y}_\tau)$. Following multi-turn DAPO (Yu et al., 2025), we treat each multi-round trajectory as a single rollout. By computing a group-normalized advantage $\hat{A}_\tau$ from the scalar rewards and broadcasting it uniformly to all tokens across the trajectory, we update f_θ to maximize the expected return:

$$\mathcal{J}_{\text{RL}}(\theta) = \mathbb{E}_{f_\theta}\Big[\tfrac{1}{|\tau|} \textstyle\sum_{i,t} \min\big(r_{i,t}\,\hat{A}_\tau,\ \text{clip}(r_{i,t}, 1-\epsilon_l, 1+\epsilon_h)\,\hat{A}_\tau\big) \Big], \tag{3}$$

where $r_{i,t}$ is the token-level importance ratio and $|\tau|$ is the trajectory token count. The same judge is used at evaluation time. Sec. 6.1 verifies that this introduces no judge bias, via cross-judge and human-agreement analyses.

5 Experiment Results

5.1 Experimental Setup

Training Details We train memory construction models on 8 NVIDIA A800 GPUs using Qwen3-4B (Yang et al., 2025a). For SFT, we use a 1×10^{-5} learning rate with cosine scheduling, 0.05 warmup ratio, and effective batch size 32 over 1000 steps. For RL training, we use a 1×10^{-6} learning rate, 20 warmup steps, batch size 32, and 8 rollouts per batch for 500 steps. Additional hyperparameters are provided in App. D.

Implementation Details We evaluate on the RoleMemo evaluation split with 32k-token conversation histories, ensuring no persona overlap with the training set. Nine baselines are compared with their native zero-shot configurations, including a NoMem setting (context only) and eight persona-agnostic memory frameworks: HiMem (Zhang et al., 2026), O-Mem (Wang et al., 2025a), MIRIX (Wang and Chen, 2025), LightMem (Fang et al., 2025), SimpleMem (Liu et al., 2026), Mem0 (Chhikara et al., 2025), PreMem (Kim et al., 2025), and Mem-α (Wang et al., 2025d). Baselines use DeepSeek-V3.2 (685B) as the memory construction model, except Mem-α, which uses its fine-tuned checkpoint. For a fair comparison, we replace each method's retrieval component with Qwen3-Embedding-0.6B (Zhang et al., 2025c), retrieving the top 10 most relevant entries per query, which we found provides sufficient context breadth. Finally, each method employs the same role-playing agent (ByteDance, 2024) to generate responses.

Evaluation Metrics Recall@10 measures whether the top 10 retrieved entries contain query-relevant ground-truth entries, evaluated separately for factual and insight memory. For role-playing, we adapt CharacterEval and employ GPT-5.1 (OpenAI, 2025) as an LLM judge to score responses on a 5-point scale (averaged over three independent judge calls on fixed model outputs) across four dimensions: *information richness*, *logical quality*, *character consistency*, and *conversational attractiveness*. More details of metric definitions are in App. E.

Table 2: Memory construction quality (Fact and Insight Recall@10) across four task types. The symbol † denotes frameworks that impose a fixed hierarchical schema on memory organization.

Framework	Interpretive Attribution		Contradiction Revelation		Value Judgment		Decision Guidance		Average	
	Fact	Insight	Fact	Insight	Fact	Insight	Fact	Insight	Fact	Insight
Baseline Agentic Memory Framework										
Memalpha†	0.15	0.05	0.23	0.10	0.21	0.10	0.21	0.09	0.20	0.08
HiMem†	0.37	0.22	0.35	0.36	0.37	0.27	0.38	0.23	0.37	0.27
Mirix†	0.43	0.12	0.37	0.15	0.43	0.18	0.45	0.13	0.42	0.15
LightMem	0.70	0.31	0.72	0.35	0.68	0.39	0.65	0.29	0.69	0.33
O-Mem†	0.70	0.27	0.66	0.38	0.70	0.35	0.66	0.30	0.68	0.33
PreMem	0.69	0.37	0.70	0.35	0.65	0.37	0.70	0.31	0.69	0.35
SimpleMem	0.69	0.35	0.70	0.41	0.70	0.48	0.71	0.41	0.70	0.41
Mem0	0.74	0.30	0.72	0.36	0.74	0.45	0.76	0.34	0.74	0.36
DualMem Framework										
DualMem-SFT	0.76	0.69	**0.77**	0.61	0.74	0.65	0.77	0.65	0.76	0.65
DualMem-RL	**0.80**	**0.78**	0.74	**0.71**	**0.75**	**0.76**	**0.79**	**0.67**	**0.77**	**0.73**

5.2 Memory Quality Evaluation

We first examine the memory construction quality of various frameworks via two dimensions: **Fact** and **Insight**. Specifically, fact recall evaluates whether the framework has extracted the necessary information to answer a query, while insight recall assesses whether the framework has constructed deeper interpretations from this information.

Fact Extraction Quality Tab. 2 reveals that memory architecture significantly affects fact recall. Hierarchical frameworks, including HiMem and Mirix, achieve only 0.37 to 0.42 average recall, as their fixed-schema structures constrain retention of fine-grained facts. O-Mem reaches 0.68, while non-hierarchical frameworks attain 0.69 to 0.74. DualMem-RL and DualMem-SFT reach 0.77 and 0.76, exceeding all DeepSeek-V3.2-driven baselines. This fact extraction ability also generalizes beyond RoleMemo. On LoCoMo (Maharana et al., 2024), DualMem-RL improves over the untrained Qwen3-4B, confirming that RoleMemo training transfers to out-of-distribution factual recall, as detailed in App. F.

Insight Construction Quality Despite high fact recall, all baselines struggle with insight construction. The best baseline reaches only 0.41 average insight recall, suggesting these frameworks do not generalize to cross-event persona-driven reasoning. In contrast, DualMem-RL reaches 0.73 and DualMem-SFT achieves 0.65, both above all baselines. To complement this matching-based metric, we score insight content with an independent LLM judge that does not observe RoleMemo annotations (App. G). Both DualMem variants produce high-quality insights approaching RoleMemo's ground-truth under independent assessment. Moreover, RL training outperforms SFT on both insight recall and LLM-judge quality, consistent with trajectory-level optimization generalizing beyond individual examples.

Task-Specific Performance Analysis Performance varies across task types, with *Decision Guidance* exhibiting the lowest insight recall across all frameworks. For instance, DualMem-RL achieves 0.73 insight recall on average but only 0.67 on *Decision Guidance*. We attribute this gap to construction difficulty rather than retrieval: fact recall remains comparable across tasks, but *Decision Guidance* requires inferring character choices from facts with indirect thematic relevance, making insight synthesis harder.

5.3 Role-playing Quality Evaluation

Having established that our approach constructs higher-quality memory, we now evaluate its downstream impact on role-playing performance.

Overall Performance and Dimension Analysis As shown in Tab. 3, despite top-performing baselines achieving fact recall comparable to our method, existing frameworks fail to translate memory quality into effective role-playing performance. This gap stems from the absence of interpretive guidance: without it, the role-playing agent cannot discern how to utilize retrieved facts from the persona's perspective, leaving them underutilized. By contrast, our framework constructs insight cognition that provides directional context, enabling the agent to synthesize retrieved facts into character-consistent responses. Consequently, DualMem-RL achieves 4.22 *information richness* and 3.78

logical quality, outperforming the strongest baseline Mem0 by 0.12 and 0.23 points, respectively. This advantage extends to persona alignment, with 4.37 *character consistency* and 4.27 *conversational attractiveness*, yielding a 4.16 overall average that surpasses the 3.94 to 4.01 range of persona-agnostic frameworks. The ranking holds on a held-out subset of 200 queries regenerated by Claude-Sonnet-4.6, suggesting that the advantage stems from persona-conditioned interpretation rather than generator-specific style (App. H).

Table 3: Role-playing quality evaluation. **Info.**, **Logic**, **Consis.**, and **Attr.** denote information richness, logical quality, character consistency, and conversational attractiveness.

Framework	Info.	Logic	Consis.	Attr.	Average
Base Setting					
NoMem	3.65	3.29	4.08	4.02	3.76
Baseline Agentic Memory Framework					
Memalpha†	4.05	3.51	4.15	4.08	3.95
HiMem†	4.05	3.49	4.13	4.08	3.94
Mirix†	4.06	3.54	4.16	4.13	4.01
LightMem	4.07	3.55	4.19	4.13	4.00
O-Mem†	4.07	3.54	4.17	4.12	3.98
PreMem	4.09	3.56	4.18	4.13	3.99
SimpleMem	4.10	3.53	4.15	4.13	3.97
Mem0	4.10	3.55	4.18	4.15	4.00
DualMem Framework					
DualMem-SFT	4.18	3.77	4.35	4.26	4.15
DualMem-RL	**4.22**	**3.78**	**4.37**	**4.27**	**4.16**
⌞ *w/o Insight*	4.12	3.57	4.19	4.16	4.01
⌞ *w/o Fact*	3.98	3.64	4.25	4.15	4.00
⌞ *w/o Training*	4.12	3.62	4.12	4.19	4.02

Ablation Analysis Tab. 3 also presents ablation variants measuring the contribution of training and each memory component. Removing RoleMemo training (*w/o Training*) while retaining the dual-stream architecture yields only 4.02 on average, confirming that training rather than architecture alone drives the gains. Among trained variants, removing insight cognition (*w/o Insight*) drops *logical quality* from 3.78 to 3.57, within 0.02 of Mem0's 3.55, confirming that inference-time reinterpretation is unreliable without stored insight (He et al., 2025; Chen et al., 2026). Removing fact cognition (*w/o Fact*) reduces *information richness* from 4.22 to 3.98, with *character consistency* dropping modestly to 4.25 as insight partially compensates. Together, training, fact cognition, and insight cognition each contribute independently to high-fidelity role-playing.

Table 4: Pearson correlation coefficients between human annotators and the LLM judge.

	Human$_1$	**Human$_2$**	**LLM Judge**
Human$_1$	–	0.88	0.82
Human$_2$	0.88	–	0.84
LLM Judge	0.82	0.84	–

Table 5: Cross-judge stability between GPT-5.1 and Gemini-3-Pro, measured as mean relative score deviation across four dimensions.

Framework	Judge	Info.	Logic	Consis.	Attr.	Agreement
NoMem	GPT5.1	3.65	3.29	4.08	4.02	99.32%
	Gemini-3-Pro	3.68	3.32	4.10	4.04	
O-Mem	GPT5.1	4.07	3.54	4.17	4.12	99.29%
	Gemini-3-Pro	4.09	3.58	4.19	4.15	
DualMem-RL	GPT5.1	4.22	3.78	4.37	4.27	99.28%
	Gemini-3-Pro	4.25	3.81	4.39	4.31	

6 Analysis

6.1 Reliability of LLM-based Evaluation

We assess the validity of LLM-based evaluation on two dimensions: human-LLM agreement and cross-judge stability. This involves 200 randomly sampled queries from the evaluation set, with detailed protocols in App. I.

Tab. 4 reports Pearson correlations among two human annotators and GPT-5.1 judge. The results show inter-annotator agreement ($r = 0.88$) and correlation between each annotator and the LLM judge ($r = 0.82$ and 0.84), indicating that GPT-5.1 scores track human judgment rather than reflecting optimization toward a specific reward signal.

To further verify stability across different judge models, Tab. 5 compares GPT-5.1 scores against Gemini-3-Pro (Google DeepMind, 2025). Agreement rates exceed 99% (defined as mean relative score deviation; App. I), with absolute score differences below 0.04 across all dimensions. The relative performance hierarchy remains consistent across both judges. DualMem-RL maintains its ranking advantage across all baselines and dimensions, indicating that the evaluation results are stable across judge models.

6.2 Transferability of Insight Construction

We investigate whether existing frameworks can reproduce these persona-task gains via prompt engineering alone. We experiment on O-Mem and PreMem, two frameworks using DeepSeek-V3.2 whose memory pipelines require no structural modification. As shown in Tab. 7, incorporating insights yields 0.09 (O-Mem*) and 0.05 (PreMem*) point improvements, yet these gains are smaller

Table 6: Qualitative comparison in a psychological counseling scenario. All responses are generated by the same role-playing agent conditioned on memory retrieved from each framework.

GT Memory	**Facts:** Ms. Zhang missed choir due to tuberculosis treatment; neighbors have been avoiding her lately. **Insight:** Social withdrawal stems from fear of stigma, not physical discomfort.
Query	Why is Ms. Zhang reluctant to participate in community activities lately?
Role-playing Agent	
∟ *with NoMem*	She might feel lonely. For example, she talks to herself at the TV.
∟ *with O-Mem*	She might still be recovering from tuberculosis. We should check if her symptoms have improved.
∟ *with* DUALMEM-SFT	She's probably feeling down after being sick for so long. Let's check if she's developing depression and might need medication.
∟ *with* DUALMEM-RL	She's probably worried about how people see her now. Let's talk about how it felt when neighbors started avoiding her.

than those of our 4B model trained on RoleMemo. Moreover, as shown in App. J, adding the insight slot does not degrade baselines' factual recall on the external benchmark LoCoMo, confirming that insight interpretation adds capability without sacrificing factual performance.

Table 7: Performance gains from incorporating persona-driven insights into existing frameworks. Methods with * use modified prompts to generate insights.

Framework	Info.	Logic	Consis.	Attr.	Average
O-Mem	4.07	3.54	4.17	4.12	3.98
O-Mem*	4.15 (+0.08)	3.67 (+0.13)	4.29 (+0.12)	4.17 (+0.05)	4.07 (+0.09)
PreMem	4.09	3.56	4.18	4.13	3.99
PreMem*	4.11 (+0.02)	3.61 (+0.05)	4.25 (+0.07)	4.17 (+0.04)	4.04 (+0.05)

6.3 Robustness Across Context Lengths

To evaluate the robustness of our framework under extended interactions, we scale the context length from 32k to 256k tokens. Fig. 3 illustrates the performance of different frameworks across varying context lengths. As context increases, all persona-agnostic baseline frameworks exhibit a significant performance decline. This suggests that accumulated noise progressively degrades the quality of extracted memories, hindering effective utilization by the role-playing agent.

In contrast, our DUALMEM-RL demonstrates remarkable stability across all scales, maintaining a high average score of 4.12 even at 256k tokens. However, DUALMEM-SFT gradually declines, primarily because SFT optimization lacks the end-to-end trajectory learning that enables RL to maintain interpretive quality under noisy conditions. This underscores the importance of RL training for robust long-context memory construction.

6.4 Case Study

Tab. 6 presents a psychological counseling case that illustrates how different memory frameworks support persona-driven conversation. When queried about recent social withdrawal, the role-playing

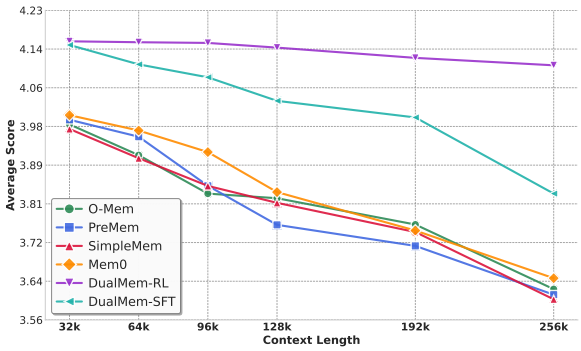

Figure 3: Performance across conversation lengths from 32k to 256k tokens. Our frameworks maintain stability while baselines degrade with increasing context.

agent without memory generates fabricated details due to absent historical context. O-Mem retrieves relevant factual evidence but lacks persona-driven insight as guidance, leaving the agent unable to generate high-fidelity responses.

Conversely, DUALMEM-RL recognizes social avoidance patterns, applying anticipated stigma insights to diagnose underlying fear of judgment. Despite generating insights, DUALMEM-SFT attributes withdrawal to prolonged illness rather than social stigma, reflecting misaligned interpretation. We further examine the limitations of DUALMEM-RL in App. K.

7 Conclusion

Existing agentic memory frameworks adopt persona-agnostic summarization, failing to provide the interpretive insight necessary for effective role-playing. We introduce RoleMemo, a dataset requiring persona-driven interpretation over conversation histories, revealing this limitation. To address this gap, we propose a dual memory framework constructing factual and insight cognition through specialized training. Results show our approach improves role-playing quality, surpassing persona-agnostic baselines. Future work will explore scaling to richer, longer contexts.

Limitations

While RoleMemo advances persona-driven memory construction, several limitations remain. First, our evaluation extends conversation histories to 256k tokens, yet real-world applications may involve multi-million token histories requiring more efficient incremental processing. Second, our training encompasses 2,052 personas across 23 domains, but generalization to unbounded real-world persona diversity remains untested in naturalistic deployment. Moreover, since RoleMemo is generated by a single LLM, the dataset may encode that generator's particular style of persona reasoning, and performance on RoleMemo may partially reflect alignment with the generator rather than a fully model-agnostic notion of persona-driven interpretation. Third, RL training via DAPO (Yu et al., 2025) is computationally intensive, which may limit accessibility for resource-constrained researchers. Fourth, our evaluation relies primarily on LLM-as-judge with subset human validation. Longitudinal user studies would provide complementary evidence for sustained persona fidelity. Finally, RoleMemo focuses on four cognitive reasoning tasks and does not exhaustively cover the full spectrum of companionship capabilities such as emotional regulation or multimodal interaction.

Ethics Statement

RoleMemo uses entirely synthetic data generated by DeepSeek-V3.2 without real user information, with automated safety filtering to eliminate violence and privacy violations. However, we emphasize that systems trained on this dataset should not replace professional services in sensitive domains such as mental health counseling, and deployment in clinical settings requires rigorous validation and human oversight. We also caution that persona-driven insight generation may produce confident psychological or interpretive statements without sufficient evidential grounding; downstream applications should surface uncertainty and avoid presenting model inferences as authoritative judgments. The dataset may inadvertently encode stereotypes from the underlying LLM despite our quality controls. Training requires significant computational resources, contributing to environmental impact. To promote responsible research, we release complete training details, prompts, and evaluation protocols in the Appendix, and plan to open-source the dataset and code under permissive licenses with usage guidelines emphasizing research purposes.

Acknowledgments

We thank the anonymous reviewers and action editors for their constructive feedback. This work was sponsored by the the National Natural Science Foundation of China under Grant No. U24A20326 and CCF-NetEase ThunderFire Innovation Research Funding (No. CCF-Netease 202601).

References

Junshuai An, Jianping Hu, Guozhu Zhang, and Dongtang Ma. 2026. Route optimization based on quantum-weighted long-and short-term memory networks. *Chinese Journal of Electronics*, 35(1):196–204.

Anthropic. 2026. Claude Sonnet 4.6. https://www.anthropic.com/claude. Accessed: 2026-08-25.

Frederic Charles Bartlett. 1995. *Remembering: A study in experimental and social psychology*. Cambridge university press.

ByteDance. 2024. Doubao: AI character chat. https://www.doubao.com. Accessed: 2026-03-13.

Hongzhan Chen, Hehong Chen, Ming Yan, Wenshen Xu, Gao Xing, Weizhou Shen, Xiaojun Quan, Chenliang Li, Ji Zhang, and Fei Huang. 2024. SocialBench: Sociality evaluation of role-playing conversational agents. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 2108–2126.

Tiantian Chen, Jiaqi Lu, Ying Shen, and Lin Zhang. 2026. ES-MemEval: Benchmarking conversational agents on personalized long-term emotional support. In *Proceedings of the ACM Web Conference 2026*, pages 5810–5821.

Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. 2025. Mem0: Building production-ready AI agents with scalable long-term memory. In *Proceedings of the 28th European Conference on Artificial Intelligence (ECAI 2025)*, volume 413 of *Frontiers in Artificial Intelligence and Applications*. IOS Press.

Nicholas Epley, Boaz Keysar, Leaf Van Boven, and Thomas Gilovich. 2004. Perspective taking as egocentric anchoring and adjustment. *Journal of personality and social psychology*, 87(3):327.

Jizhan Fang, Xinle Deng, Haoming Xu, Ziyan Jiang, Yuqi Tang, Ziwen Xu, Shumin Deng, Yunzhi Yao, Mengru Wang, Shuofei Qiao, and 1 others. 2025. LightMem: Lightweight and efficient memory-augmented generation. *arXiv preprint arXiv:2510.18866*.

Google DeepMind. 2025. Gemini 3 Pro model card. https://deepmind.google/models/model-cards/gemini-3-pro/. Accessed: 2026-08-25.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, and 1 others. 2025. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning. *Nature*, 645(8081):633–638.

Junqing He, Liang Zhu, Rui Wang, Xi Wang, Gholamreza Haffari, and Jiaxing Zhang. 2025. MADial-Bench: Towards real-world evaluation of memory-augmented dialogue generation. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 9902–9921.

Ruofan Hu, Yan Xia, Minjie Hong, Jieming Zhu, Bo Chen, Xiaoda Yang, Minghui Fang, and Tao Jin. 2025a. Vela: scalable embeddings with voice large language models for multimodal retrieval. *arXiv preprint arXiv:2506.14445*.

Ruofan Hu, Menghui Zhu, Jieming Zhu, Bo Chen, Shengyang Xu, Minjie Hong, Xiaoda Yang, Sashuai Zhou, Li Tang, Tao Jin, and 1 others. 2026. Docretriever: A plug-and-play framework for multimodal document retrieval with comprehensive benchmark. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 1780–1791.

Yuanzhe Hu, Yu Wang, and Julian McAuley. 2025b. Evaluating memory in LLM agents via incremental multi-turn interactions. *arXiv preprint arXiv:2507.05257*.

Bowen Jiang, Zhuoqun Hao, Young-Min Cho, Bryan Li, Yuan Yuan, Sihao Chen, Lyle Ungar, Camillo J. Taylor, and Dan Roth. 2025a. Know me, respond to me: Benchmarking LLMs for dynamic user profiling and personalized responses at scale. *arXiv preprint arXiv:2504.14225*.

Bowen Jiang, Yuan Yuan, Maohao Shen, Zhuoqun Hao, Zhangchen Xu, Zichen Chen, Ziyi Liu, Anvesh Rao Vijjini, Jiashu He, Hanchao Yu, and 1 others. 2025b. PersonaMem-v2: Towards personalized intelligence via learning implicit user personas and agentic memory. *arXiv preprint arXiv:2512.06688*.

Sangyeop Kim, Yohan Lee, Sanghwa Kim, Hyunjong Kim, and Sungzoon Cho. 2025. PreMem: Pre-storage reasoning for episodic memory: Shifting inference burden to memory for personalized dialogue. *arXiv preprint arXiv:2509.10852*.

Aixin Liu, Aoxue Mei, Bangcai Lin, Bing Xue, Bingxuan Wang, Bingzheng Xu, Bochao Wu, Bowei Zhang, Chaofan Lin, Chen Dong, and 1 others. 2025. DeepSeek-V3.2: Pushing the frontier of open large language models. *arXiv preprint arXiv:2512.02556*.

Jiaqi Liu, Yaofeng Su, Peng Xia, Siwei Han, Zeyu Zheng, Cihang Xie, Mingyu Ding, and Huaxiu Yao. 2026. SimpleMem: Efficient lifelong memory for LLM agents. *arXiv preprint arXiv:2601.02553*.

Adyasha Maharana, Dong-Ho Lee, Sergey Tulyakov, Mohit Bansal, Francesco Barbieri, and Yuwei Fang. 2024. Evaluating very long-term conversational memory of LLM agents. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13851–13870.

Elliot Nelson, Georgios Kollias, Payel Das, Subhajit Chaudhury, and Soham Dan. 2024. Needle in the haystack for memory based large language models. *arXiv preprint arXiv:2407.01437*.

OpenAI. 2025. GPT-5.1: A smarter, more conversational ChatGPT. https://openai.com/index/gpt-5-1/. Accessed: 2026-08-25.

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.

Egor Pakhomov, Erik Nijkamp, and Caiming Xiong. 2025. ConvoMem benchmark: Why your first 150 conversations dont need RAG. *arXiv preprint arXiv:2511.10523*.

Jeiyoon Park, Yongshin Han, Minseop Kim, and Kisu Yang. 2025. Dynamic context adaptation for consistent role-playing agents with retrieval-augmented generations. *arXiv preprint arXiv:2508.02016*.

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.

Jiahao Qiu, Xuan Qi, Hongru Wang, Xinzhe Juan, Yimin Wang, Zelin Zhao, Jiayi Geng, Jiacheng Guo, Peihang Li, Jingzhe Shi, and 1 others. 2025. Alita-G: Self-evolving generative agent for agent generation. *arXiv preprint arXiv:2510.23601*.

Shalom H Schwartz. 1992. Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology*, volume 25, pages 1–65. Elsevier.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. Reflexion: Language agents with verbal reinforcement learning. *Advances in neural information processing systems*, 36:8634–8652.

Quan Tu, Shilong Fan, Zihang Tian, Tianhao Shen, Shuo Shang, Xin Gao, and Rui Yan. 2024. CharacterEval: A Chinese benchmark for role-playing conversational agent evaluation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11836–11850.

Piaohong Wang, Motong Tian, Jiaxian Li, Yuan Liang, Yuqing Wang, Qianben Chen, Tiannan Wang, Zhicong Lu, Jiawei Ma, Yuchen Eleanor Jiang, and 1 others. 2025a. O-Mem: Omni memory system for personalized, long horizon, self-evolving agents. *arXiv preprint arXiv:2511.13593*.

Xintao Wang, Heng Wang, Yifei Zhang, Xinfeng Yuan, Rui Xu, Jen-tse Huang, Siyu Yuan, Haoran Guo, Jiangjie Chen, Shuchang Zhou, and 1 others. 2025b. CoSER: Coordinating LLM-based persona simulation of established roles. In *Forty-second International Conference on Machine Learning*.

Yongjie Wang, Jonathan Leung, and Zhiqi Shen. 2025c. RoleRAG: Enhancing LLM role-playing via graph guided retrieval. *arXiv preprint arXiv:2505.18541*.

Yu Wang and Xi Chen. 2025. MIRIX: Multi-agent memory system for LLM-based agents. *arXiv preprint arXiv:2507.07957*.

Yu Wang, Ryuichi Takanobu, Zhiqi Liang, Yuzhen Mao, Yuanzhe Hu, Julian McAuley, and Xiaojian Wu. 2025d. Mem-α: Learning memory construction via reinforcement learning. *arXiv preprint arXiv:2509.25911*.

Di Wu, Hongwei Wang, Wenhao Yu, Yuwei Zhang, Kai-Wei Chang, and Dong Yu. 2025a. LongMemEval: Benchmarking chat assistants on long-term interactive memory. In *The Thirteenth International Conference on Learning Representations*.

Xixi Wu, Kuan Li, Yida Zhao, Liwen Zhang, Litu Ou, Huifeng Yin, Zhongwang Zhang, Xinmiao Yu, Dingchu Zhang, Yong Jiang, and 1 others. 2025b. ReSum: Unlocking long-horizon search intelligence via context summarization. *arXiv preprint arXiv:2509.13313*.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025a. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

Diji Yang, Linda Zeng, Jinmeng Rao, and Yi Zhang. 2025b. Knowing you dont know: Learning when to continue search in multi-round RAG through self-practicing. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1305–1315.

Hongli Yu, Tinghong Chen, Jiangtao Feng, Jiangjie Chen, Weinan Dai, Qiying Yu, Ya-Qin Zhang, Wei-Ying Ma, Jingjing Liu, Mingxuan Wang, and 1 others. 2026. MemAgent: Reshaping long-context LLM with multi-conv RL-based memory agent. In *The Fourteenth International Conference on Learning Representations*.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Lingjun Liu, and 1 others. 2025. DAPO: An open-source LLM reinforcement learning system at scale. *arXiv preprint arXiv:2503.14476*.

Saber Zerhoudi and Michael Granitzer. 2024. PersonaRAG: Enhancing retrieval-augmented generation systems with user-centric agents. *arXiv preprint arXiv:2407.09394*.

Yunpeng Zhai, Shuchang Tao, Cheng Chen, Anni Zou, Ziqian Chen, Qingxu Fu, Shinji Mai, Li Yu, Jiaji Deng, Zouying Cao, and 1 others. 2025. AgentEvolver: Towards efficient self-evolving agent system. *arXiv preprint arXiv:2511.10395*.

Haonan Zhang, Run Luo, Xiong Liu, Yuchuan Wu, Ting-En Lin, Pengpeng Zeng, Qiang Qu, Feiteng Fang, Min Yang, Lianli Gao, and 1 others. 2025a. OmniCharacter: Towards immersive role-playing agents with seamless speech-language personality interaction. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 26318–26331.

Ningning Zhang, Xingxing Yang, Zhizhong Tan, Weiping Deng, and Wenyong Wang. 2026. HiMem: Hierarchical long-term memory for LLM long-horizon agents. *arXiv preprint arXiv:2601.06377*.

Xinliang Frederick Zhang, Nick Beauchamp, and Lu Wang. 2025b. PRIME: Large language model personalization with cognitive dual-memory and personalized thought process. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 33695–33724.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, and 1 others. 2025c. Qwen3 embedding: Advancing text embedding and reranking through foundation models. *arXiv preprint arXiv:2506.05176*.

Siyan Zhao, Mingyi Nguyen, and Aditya Grover. 2025. Do LLMs recognize your preferences? evaluating personalized preference following in LLMs. In *The Thirteenth International Conference on Learning Representations*.

Qinghua Zheng, Huan Liu, Xiaoqing Zhang, Caixia Yan, Xiangyong Cao, Tieliang Gong, Yong-Jin Liu, Bin Shi, Zhen Peng, Xiaocen Fan, and 1 others. 2025. Machine memory intelligence: Inspired by human memory mechanisms. *Engineering*.

Jinfeng Zhou, Yongkang Huang, Bosi Wen, Guanqun Bi, Yuxuan Chen, Pei Ke, Zhuang Chen, Xiyao Xiao, Libiao Peng, Kuntian Tang, and 1 others. 2025. CharacterBench: Benchmarking character customization of large language models. In *Proceedings of the AAAI*

Conference on Artificial Intelligence, volume 39, pages 26101–26110.

Appendix Roadmap

- **App. A — Relationship to Reflexion.** Differentiates DUALMEM from Reflexion along task setting, signal source, and training paradigm.
- **App. B — RoleMemo Construction Details.** Generation hyperparameters and prompts for persona profiles, insights, fact–insight pairs, and dialogue blocks.
- **App. C — Persona Domains and Schema.** 23 social-thematic domains, five value-orientation axes, and full persona schema.
- **App. D — Training Hyperparameters.** SFT and DAPO hyperparameters not reported in the main text.
- **App. E — Evaluation Metrics.** Semantic matching protocol for Recall@10 ($\tau = 0.7$, stable for $\tau \in [0.65, 0.75]$) and four role-playing quality dimension definitions.
- **App. F — Generalization to LoCoMo.** RoleMemo training transfers to out-of-distribution factual recall without overfitting.
- **App. G — Insight Quality under an Independent Judge.** Blind GPT-5.1 judge confirms DUALMEM insights are not stylistic artifacts of the training distribution (4.62 vs. ground-truth 4.67).
- **App. H — Cross-Generator Robustness.** Ranking advantage of DUALMEM-RL holds on a Claude-Sonnet-4.6-regenerated subset, ruling out generator-specific style as a confound.
- **App. I — LLM Judge Reliability Protocol.** Human–LLM agreement ($r = 0.82$–0.84) and cross-judge stability (GPT-5.1 vs. Gemini-3-Pro, $> 99\%$) validate automated evaluation.
- **App. J — No-Regression on LoCoMo.** Adding an insight slot does not reduce baseline factual recall; insight cognition is an additive capability.
- **App. K — Failure Mode Analysis.** Three recurring failure patterns in DUALMEM-RL: insight not recognized, facts ignored behind adopted insight, and surface-similarity conflation.
- **App. L — Human Evaluation Details.** Annotator recruitment, compensation (100 RMB/hr), and privacy protocols.

A Relationship to Reflexion

This section addresses concerns about the differentiation between DUALMEM's factual–insight architecture and Reflexion's observation–reflection pattern. While both adopt a two-level structure of base evidence plus a higher-order representation, the two frameworks differ along three dimensions.

- **Task setting.** Reflexion targets task-oriented reasoning (code generation, decision-making, QA) where success is objectively verifiable. DUALMEM targets persona-driven role-playing dialogue, where the goal is high-fidelity character expression rather than task completion.
- **Signal source.** Reflexion's reflection is reactive: it is triggered by an external signal and reasons about what went wrong. DUALMEM's insight cognition has no such failure signal at construction time. Insight is generated proactively from new facts through the persona lens, regardless of any downstream outcome. Reflexion reacts to failure; DUALMEM proactively interprets evidence.
- **Training paradigm.** Reflexion is an inference-time prompting framework with no parameter update. DUALMEM is trained end-to-end on RoleMemo via SFT and multi-turn trajectory RL (Sec. 4.3), internalizing persona-driven interpretation into model weights rather than relying on prompt scaffolding.

More broadly, the observation–reflection structure has become a shared paradigm across modern agent systems, independently adopted by Generative Agents (Park et al., 2023), AgentEvolver (Zhai et al., 2025), ReSum (Wu et al., 2025b), and Alita-G (Qiu et al., 2025), each in a different domain. DUALMEM's contribution lies within this paradigm: it is the first persona-driven instantiation, with persona-conditioned insight as the reflection layer, validated by a long-context role-playing benchmark and a trained 4B specialist.

B RoleMemo Construction Details

B.1 Generation Hyperparameters

All dataset instances are generated using DeepSeek-V3.2 via the OpenAI-compatible API; full parameters are listed in Tab. 8.

Table 8: DeepSeek-V3.2 Generation Parameters

Parameter	Value
Model	DeepSeek-V3.2
Temperature	0.8
Response Format	JSON Object
Stream	False

B.2 Cognitive Insight Generation Prompts

The prompt for generating 20 persona-conditioned cognitive insights is shown below.

Insight Generation Prompt

Role: High-Quality Persona Insight Generator
Task: You are an expert in psychology, sociology, and systems thinking. Generate cognitive insights aligned with the given NPC persona. These insights will serve as the foundation for designing memory evaluation queries.
Input Format: The persona description includes:

- Core values
- Personality traits
- Background and experiences
- Domain expertise / professional perspective
- Key goals or conflicts
- Work and hobbies

Insight Generation Principles:

- **Perspective Uniqueness:** Must reflect the persona's unique cognitive framework, not common sense.
- **Verifiability and Specificity:** Insights must be derivable from specific user behaviors, dialogues, or event combinations.
- **Counter-Intuitive Potential:** Ideal insights exhibit counter-intuitive tension between surface phenomena and underlying truth.
- **Reasoning Complexity:** Insights should require at least two reasoning steps.
- **Fact Combination Necessity:** Each insight should require at least two different types of factual fragments for reliable derivation.

Insight Categories:

- **[Decision Guidance]:** Provides concrete, actionable behavioral guidance based on the persona's unique value hierarchy.
- **[Interpretive Attribution]:** Explains deep, systemic causes behind phenomena, beyond surface-level explanations.
- **[Value Judgment]:** Evaluates things, behaviors, or states based on the persona's unique value system.
- **[Contradiction Revelation]:** Reveals systemic contradictions between surface phenomena and underlying logic.

[Detailed application scope, generation workflow, and examples are omitted for brevity.]
Output Format:

```
{
  "role_description": "Brief
      summary of the input persona",
  "core_values": ["Core value 1", "
      Core value 2"],
  "insights": [
    "[Category]: Specific insight
        description 1",
    "[Category]: Specific insight
        description 2",
    "..."
  ]
}
```

B.3 Fact-Insight Pair and Query Generation

The prompt for generating factual fragments and queries from each insight is shown below.

Fact-Insight Pair Generation Prompt

Role: Persona-Driven Memory Test Designer
Background: Generic memory models only record facts, while persona-driven memory models extract deep interpretations (insights) from facts based on specific values and reasoning patterns. This designer generates challenging test cases by examining whether models can derive key insights from two seemingly mundane facts aligned with the user's identity.
Task: Given the [Persona Insight], [NPC Persona], [User Profile], and [Reasoning Hints], design the following three components:

1. Two factual fragments that constitute the persona insight: user-mentioned personal experiences, observations, industry con-

cerns, personal reflections, or media exposure that align with their identity.

2. A test question that requires the insight to interpret these two facts for the optimal answer. May incorporate reasoning hints or be entirely self-designed.
3. A logic connector explaining how fragments → insight → question are linked.

Constraints & Rules (Core Requirements):

- **Identity Anchoring:** Facts must align with [User Profile] and occur across diverse settings.
- **Factual Fragment Design:** Facts must be neutral, mundane, and superficially irrelevant. Their combination, when interpreted through the persona's reasoning, yields a notably stronger and more coherent insight than literal retrieval alone.
- **Test Question Design:** The question must not directly use core concepts from the insight, must explicitly rely on specific details from factual fragments, and should reward viewing both events through the insight with a more complete and persona-aligned response, while still admitting literal answers from facts alone.

[Detailed examples, negative case prevention guidelines, and workflow are omitted for brevity.]

Output Format:

```
{
  "insight": "The given persona
      insight",
  "fragments": [
    "Factual fragment 1: A neutral,
         mundane user statement",
    "Factual fragment 2: Another
        neutral, mundane user
        statement"
  ],
  "test_question": {
    "text": "A natural question
        that doesn't directly hint
        at the insight and requires
        facts for a good answer",
    "type": "Decision/Attribution/
        Value Judgment/Contradiction
         Resolution"
  },
  "connector": "Concise explanation
      of the logic chain"
}
```

B.4 Conversational Block Generation

The prompt for weaving facts into 24-turn dialogue blocks is shown below.

Natural Conversation Generation Prompt

Role: Natural Dialogue Generator (Topic-Anchored Deep Conversation)

Core Principle: You must treat the casual topic as the sole backbone and core content of the conversation, expanding extensively around the topic's multiple sub-dimensions (e.g., if topic is "urban landmarks and cultural connotations," extend to architectural styles, historical background, photo spots, cultural symbols). Factual fragments are merely incidental memories triggered when User discusses the topic—like glancing at an old poster while walking, mentioned and passed over. They should be naturally embedded without warranting targeted responses or extensions.

Core Task: Given a casual topic, User persona, NPC persona, and two factual fragments, generate approximately 24 rounds (48 messages, User and NPC alternating) of natural deep conversation. The dialogue must fully expand around the topic's multiple sub-dimensions, with each turn tightly focused on topic-related details to achieve "thorough topic exploration." Individual utterances should be concise (1-2 sentences primarily).

Generation Rules (Strict Compliance):

- **Topic Priority with Deep Extension:** Dialogue must revolve around the casual topic, with each utterance containing specific words directly related to the topic.
- **Fragment Embedding (Four-No Principles):**
  - *No awkwardness:* Fragments serve as background, merged into topic-related actions/scenarios.
  - *No independence:* Fragments must appear as subordinate clauses or parentheticals attached to topic-core main clauses.
  - *No evaluation:* Objectively state events without adding subjective evaluations.
  - *No response:* NPC absolutely must not extend, question, or comment on fragment content.

- **NPC Behavior:** Every NPC response must focus on the topic and advance topic depth.

[Detailed case studies and generation steps are omitted for brevity.]

Input Format:

```
{
  "user_persona": "User persona
      description (personality,
      occupation, speaking style)",
  "npc_persona": "NPC persona
      description (personality,
      occupation, speaking style)",
  "talk_topics": "Main casual topic
       (e.g., 'urban landmarks', '
      college life', 'recent meals')
      ",
  "fragments": ["Factual fragment 1
      ", "Factual fragment 2"]
}
```

Output Format:

```
{
  "dialogue": [
    {"role": "User", "content": "
        User statement (with topic-
        specific words)"},
    {"role": "Assistant", "content"
        : "NPC statement (with topic
        -specific words, advancing
        topic)"},
    // ... approximately 24 rounds
        (48 messages, User and
        Assistant alternating)
  ],
  "fragment_positions": {
    "fragment_1": "Complete User
        statement containing
        fragment 1",
    "fragment_2": "Complete User
        statement containing
        fragment 2"
  }
}
```

C Persona Domains and Schema

C.1 Domain Organization

RoleMemo covers 23 social-thematic domains, organized along five value-orientation axes adapted from Schwartz's basic human values framework (Schwartz, 1992). Tab. 9 lists the five groups.

C.2 Persona Schema

Each persona is a structured record over the fields in Tab. 10, and all fields are strings.

Fig. 4 shows a representative persona from the Public-Interest Law domain, where lifestyle and derived fields are abbreviated for space.

Table 9: Five-group organization of the 23 RoleMemo domains, aligned with axes of Schwartz's human values.

Group	Value prior	Domains
G1. Rights and Equity	universalism, protective fairness	Public-Interest Law; Educational Equity; Disability Vocational Empowerment; Mental-Health and Vulnerable-Group Services; Public-Health Service
G2. Community and Governance	conservation, benevolence	Community Cultural Development; Community Governance and Spatial Optimization; Urban Micro-Renewal; Rural Development and Grassroots Construction; Conflict Transformation and Collaboration
G3. Culture, Heritage, and Identity	tradition vs. self-direction	Cultural Heritage Preservation; Traditional Culture Innovation; Minority Craft Inheritance; Niche and Subculture Scenes; Humanistic Observation and Social Research
G4. Wellbeing and Sustainability	stewardship-oriented universalism	Healthy Living and Psychological Support; Silver Economy; Environment and Sustainability; Stray Animal Protection and Ecological Harmony
G5. Frontier Work and Tech Ethics	openness with responsibility	Technology Ethics and Digital Civilization; Workplace and Personal Growth; Workplace Ecology and Vocational Empowerment; Educational and Public-Service Innovation

C.3 Construction Pipeline

Stage 1: Seed scaffolding Authors fix the 23 domains and seed `titles`; an LLM expands the seed set, and authors screen every additional `title` before generation.

Stage 2: LLM expansion For each retained `title`, we prompt DeepSeek-V3.2 (temperature 0.8, JSON-object response) to expand a full persona record over the fields in Tab. 10 (except `content`, which is a deterministic concatenation). The prompt is conditioned on the domain and `title`, and generates an internally consistent (`core_values`, `personality`, `background`, `dilemma`) tuple.

Stage 3: Random-sample human curation The authors manually audit randomly sampled personas using a four-part rubric: **internal coherence** (no contradiction among value, personality, background, and dilemma), **domain fit** (`title` and `current_status` belong to the declared domain), **value anchoring** (a recognizable value prior is present; neutral AI-style personas are rejected),

Table 10: Persona schema. The relational dimension is carried implicitly by `dilemma` and `current_status`.

Dimension	Field	Role
meta	`id`, `category`	identifier and sub-theme label
identity	`role_name`	persona's name
identity	`title`	occupational or social role
identity	`background`	formative experience leading to the role
identity	`current_status`	current affiliation and focus areas
identity	`hometown`	geographic grounding
value	`core_values`	value prior that drives interpretation
value	`goals`	persona-level objectives
value	`dilemma`	recurring friction (implicit relational cue)
behavioral	`personality`	overt and latent traits
behavioral	`expression_style`	register of verbal output
behavioral	`hobbies`	role-aligned professional habits
behavioral	`daily_hobby`	off-duty lifestyle habit
behavioral	`favorite_food/animal/plant`	grounding details for small talk
behavioral	`preferred_travel_city`	grounding detail for small talk
derived	`content`	concatenation used as the role-play system prompt

and **diversity** (`title` and `dilemma` angles do not collapse within a domain). Failing personas are regenerated or lightly revised. We release the rubric, the seed `title` list, and the expansion prompts alongside the dataset for reproducibility.

D Training Hyperparameters

Tab. 11 and Tab. 12 list the hyperparameters not stated in Sec. 5.1 (Experimental Setup); settings already given in the main text (base model, learning rates, schedulers, effective batch sizes, training steps) are not repeated here. vLLM rollout sampling uses the default configuration (temperature 1.0, top-p 1.0) and is omitted from the RL table.

Table 11: Additional SFT training hyperparameters.

Parameter	Value
Training Framework	LLaMA-Factory
Precision	BF16
Per-device Batch Size	4
Gradient Accumulation Steps	8
Cutoff Length	6144
Gradient Checkpointing	True
DeepSpeed Stage	3 (Offload)

Table 12: RL training hyperparameters.

Parameter	Value
Training Framework	verl
Algorithm	DAPO
PPO Mini Batch Size	4
Max Prompt Length	8192
Max Response Length	1024
KL Loss Coefficient	0.001
Entropy Coefficient	0.0
Clip Ratio (high)	0.20
Clip Ratio (low)	0.10
Loss Aggregation	token-mean

E Evaluation Metrics

E.1 Recall@10 Matching Protocol

Since memory entries from different frameworks vary in surface form especially in LLM-based retrieval (Hu et al., 2025a, 2026), we compute both Fact and Insight Recall@10 with semantic rather than exact matching. For each ground-truth entry g_i and the top-10 retrieved entries $\{r_j\}$, we encode all entries with Qwen3-Embedding-0.6B and count g_i as recalled if $\max_j \cos(\mathrm{emb}(g_i), \mathrm{emb}(r_j)) \geq 0.7$. Recall@10 is the fraction of ground-truth entries recalled, macro-averaged over queries. The threshold $\tau = 0.7$ was set via pilot inspection as the lowest value at which no semantically inequivalent pair was accepted; framework rankings in Tab. 2 are stable for $\tau \in [0.65, 0.75]$. The same encoder, threshold, and aggregation are applied uniformly to every framework.

```
{
"id": 13001,
"role_name": "Zhang Shumin",
"title": "Legal-Aid Specialist for Elderly Rights",
"core_values": "Legal dignity should not decline with age; services must adapt to the cognition and mobility patterns of older adults.",
"personality": "Warm and patient when unpacking statutes; inwardly resilient, with strong empathy for the elderly.",
"background": "Witnessed a family elder fail to obtain redress after a health-product scam; moved from corporate administration to law.",
"current_status": "Deputy director of an elderly-rights legal-aid unit, focusing on property safety, consumer fraud, support disputes, and will planning.",
"goals": "Build an age-friendly legal-aid service system.",
"dilemma": "Tension between the weak evidence-awareness of older adults and the procedural rigor of rights enforcement.",
"expression_style": "Firm and grounded.",
"hometown": "Beijing",
// lifestyle fields (favorite_food, favorite_animal,
// favorite_plant, daily_hobby, preferred_travel_city)
// and derived field (content) are omitted.
}
```

Figure 4: Example persona (Public-Interest Law), translated to English. Keys are bolded. Lifestyle and derived fields are omitted for space.

E.2 Role-Playing Quality

Each of the four dimensions is scored on a 5-point Likert scale (1=Poor, 5=Excellent). The judge receives the persona profile, query, ground-truth reference response, and candidate response, then returns a score with a brief justification per dimension.

Information Richness Whether the response accurately incorporates key factual information from retrieved memory, naturally integrated into dialogue flow without forced concatenation. Responses containing irrelevant facts are penalized unless used as analogy or contextual example.

Logical Quality Whether the response contains key insights that logically address the query, with reasoning naturally woven into the reply. We strictly penalize *AI-style neutrality* (e.g., "on one hand… on the other hand… it depends on you…"), since such hedging fails to provide guidance aligned with the persona's perspective.

Character Consistency Adherence to the persona across behavior/identity alignment, utterance style and tone, and knowledge-scope adherence (no anachronistic knowledge, e.g., ancient personas referencing modern technology).

Conversational Attractiveness Engagement quality, anchored primarily by *human-likeness*: responses must sound natural, strictly penalizing "AI formatting" such as numbered lists ("1. 2. 3."), bold emphasis ("**important**"), or formulaic conclusions ("in summary…"); secondary criteria include varied, fluent expression and appropriate empathy with the user's situation.

F Generalization to LoCoMo

We test whether DualMem is overfitted to RoleMemo using LoCoMo (Maharana et al., 2024), a needle-in-a-haystack factual-recall benchmark; it cannot evaluate persona-driven interpretation and serves as an external generalization check only.

Setup We evaluate four memory-construction checkpoints: the untrained Qwen3-4B and Qwen3-8B (included as an additional size reference not used in the main experiments), our DualMem-SFT (4B), and DualMem-RL (4B), with all other components fixed. We report F1 on the four LoCoMo question types (Multi-hop, Temporal, Open-domain, Single-hop) and the Overall score.

Results Tab. 13 reports the LoCoMo scores. DualMem-RL (4B) reaches 34.0 Overall, achieving parity with the untrained Qwen3-8B (33.7) at half the parameter count, with DualMem-SFT and DualMem-RL respectively gaining +2.5 and +4.0 Overall over the same-size Qwen3-4B baseline. The largest gain appears on Multi-hop (+8.2 over Qwen3-4B), the dimension that most rewards cross-event interpretive reasoning, consistent with the design goal of DualMem.

DualMem-RL trails Qwen3-8B on Temporal by 2.7 points. RoleMemo lacks explicit temporal-ordering objectives, and we leave time-aware retrieval to future work.

Compared with DeepSeek-V3.2-driven persona-agnostic baselines on the same benchmark, our 4B DualMem-RL trails O-Mem (38.9) and PreMem (37.9) on Overall (App. J). This is expected and consistent with our scoped claim: DualMem targets persona-driven memory tasks with a small trained model, not general-purpose factual recall

with a 685B-class driver. The relevant generalization signal here is the absence of regression and the parity with Qwen3-8B shown above, not absolute headroom against much larger drivers.

G Insight Quality under an Independent Judge

The insight recall of DUALMEM in Tab. 2 could in principle reflect surface alignment with the annotation style of RoleMemo rather than the generation of high-quality persona-driven insights. We address this concern by evaluating insight content quality directly, using an independent judge that does not have access to the RoleMemo ground-truth annotations.

Setup We evaluate the insights already constructed by each method on the full RoleMemo evaluation set, without resampling. Each insight is scored independently by GPT-5.1 at temperature $= 0$. The judge receives only the persona profile, the relevant conversation excerpts, and the candidate insight. It has no access to the RoleMemo ground-truth insights and is not told which method generated the candidate.

The judge rates each insight on three dimensions on a 1–5 scale:

- **Persona Plausibility.** The insight fits this persona's identity, values, and situation, not a generic statement.
- **Evidential Grounding.** The insight is derivable from the provided conversation evidence, not unsupported or contradicted.
- **Specificity.** The insight is a substantive, non-trivial interpretation, not a platitude applicable to any persona.

We compare four methods that produce explicit insight outputs: the prompt-engineered baselines O-Mem* and PreMem* (Tab. 7), and our DUALMEM-SFT and DUALMEM-RL (other baselines in Tab. 2 produce no explicit insight field). As an upper-bound reference, we additionally score the RoleMemo ground-truth insight annotations under the same judge.

Results Tab. 14 reports the per-dimension averages. Both DUALMEM variants outperform the prompt-engineered baselines on all three dimensions; since the judge does not see RoleMemo annotations, this gap reflects the intrinsic quality of DUALMEM's insights rather than their stylistic match to the training distribution.

DUALMEM-RL approaches the ground-truth upper bound (4.62 vs. 4.67), indicating that the trained model produces insights of near-annotation quality even under independent assessment. The ground-truth annotations score 4.67, below the ceiling, consistent with the judge applying the rubric strictly rather than rewarding provenance.

H Cross-Generator Robustness

Setting We regenerate a held-out evaluation subset using Claude-Sonnet-4.6 (Anthropic, 2026) to isolate the data generator as a potential confound. Following the same protocol as App. B, we instantiate 20 new personas disjoint from both training and original evaluation splits, and produce 200 queries approximately balanced across the four task types, grounded in 32k-token conversation histories. All baselines and our DualMem variants are evaluated under the identical retriever, role-playing agent, and GPT-5.1 judge as in Tab. 3.

Result As shown in Tab. 15, the relative ranking across methods is preserved on this Claude-generated subset. DUALMEM-RL maintains a **0.20**-point overall lead over the strongest baselines (LightMem/SimpleMem at 3.95), with the insight-recall margin remaining substantial (**0.78** vs. **0.38**). Absolute scores are uniformly **0.01–0.06** points lower than the in-distribution evaluation in Tab. 3, consistent with the natural distribution shift introduced by a different generator; the uniformity of this decline across baselines and DUALMEM indicates the gap is structural rather than stylistic.

I LLM Judge Reliability Protocol

We use the same 200 queries randomly sampled from the evaluation set, with all LLM judges run at temperature$= 0$ for deterministic scoring.

Human–LLM agreement Two independent annotators score responses using identical inputs as the LLM judge. Inter-annotator and human–LLM agreement (Tab. 4) are reported as Pearson correlations between the per-query scores of each pair, averaged across the four dimensions.

Cross-judge stability We compare GPT-5.1 against Gemini-3-Pro on the same 200-query sample and report a per-framework agreement score (Tab. 5). Let $D = \{\text{Info, Logic, Consis, Attr}\}$ denote the four scoring dimensions; the agreement is

Table 13: Generalization to LoCoMo. The memory construction model is varied across four checkpoints; the retriever, role-playing agent, and judge are held fixed. Best score per column in **bold**.

Model	Multi-hop	Temporal	Open-dom.	Single-hop	Overall
Qwen3-4B	23.9	18.6	**17.9**	37.8	30.0
Qwen3-8B	28.9	**22.9**	15.3	41.6	33.7
DualMem-SFT (4B)	30.1	18.8	16.8	40.3	32.5
DualMem-RL (4B)	**32.1**	20.2	17.5	**41.8**	**34.0**

Table 14: Insight quality under an independent GPT-5.1 judge that does not see RoleMemo ground-truth annotations. **Plaus.**, **Ground.**, **Spec.** denote Persona Plausibility, Evidential Grounding, and Specificity, each scored on a 1–5 scale. The last row is the RoleMemo ground-truth as an upper-bound reference.

Method	Plaus.	Ground.	Spec.	Avg
O-Mem*	4.25	4.29	4.24	4.26
PreMem*	4.30	4.26	4.32	4.29
DualMem-SFT	4.58	4.55	4.57	4.57
DualMem-RL	4.62	4.60	4.63	**4.62**
RoleMemo GT (ref.)	4.67	4.70	4.65	4.67

defined as

$$\text{Agreement} = 1 - \frac{1}{|D|} \sum_{d \in D} \frac{|s_d^{\text{GPT}} - s_d^{\text{Gem}}|}{s_d^{\text{GPT}}},$$

defined as one minus the mean relative score deviation across the four dimensions. Across all (framework, dimension) cells, the absolute per-dimension difference $|\Delta|$ remains within 0.04.

J No-Regression of Insight Cognition on LoCoMo

We re-run O-Mem and PreMem on LoCoMo, a factual-recall benchmark, with and without an insight slot to check whether attaching insight degrades baseline factual recall. DeepSeek-V3.2 is used as the memory construction model, with the retriever, role-playing agent, and judge held fixed.

Tab. 16 reports per-dimension and Overall scores. Adding insight cognition raises O-Mem from 38.9 to **39.0** and PreMem from 37.9 to **38.1** on Overall; per-dimension shifts stay within ± 2.4 points and net out positively in both cases.

Combined with Tab. 7 in the main paper, where the same prompt-only insight extension lifted persona dialogue scores by $+0.05$ to $+0.09$, these results indicate that the insight mechanism adds capability without reducing factual performance: it brings clear gains where persona-driven interpretation is required, while leaving factual recall on neutral benchmarks intact.

Table 15: Cross-generator robustness on a subset regenerated by Claude-Sonnet-4.6. **Info.**, **Logic.**, **Consis.**, and **Attr.** denote information richness, logical quality, character consistency, and conversational attractiveness, each averaged over three GPT-5.1 judge runs.

Framework	Insight R@10	Info.	Logic	Consis.	Attr.
Base Setting					
NoMem	–	3.56	3.20	4.04	4.02
Baseline Agentic Memory Framework					
Memalpha†	0.04	3.99	3.48	4.08	3.92
HiMem†	0.14	4.05	3.54	4.08	3.96
Mirix†	0.11	4.05	3.54	4.15	3.94
LightMem	0.38	4.07	3.55	4.14	4.02
O-Mem†	0.25	4.07	3.52	4.12	3.95
PreMem	0.30	4.06	3.56	4.12	4.00
SimpleMem	0.34	4.08	3.54	4.16	4.03
Mem0	0.35	4.05	3.52	4.14	4.04
DualMem Framework					
DualMem-SFT	0.77	4.12	3.76	4.36	4.16
DualMem-RL	**0.78**	**4.24**	**3.79**	**4.40**	**4.17**

Table 16: No-regression check on LoCoMo. * denotes the variant with an insight slot added. DeepSeek-V3.2 is used as the memory construction model. Best Overall score within each baseline pair in **bold**.

Method	Multi-hop	Temporal	Open-dom.	Single-hop	Overall
O-Mem	40.3	23.9	26.4	45.6	38.9
O-Mem*	37.9	24.0	29.6	46.2	**39.0**
PreMem	35.0	27.3	29.4	44.0	37.9
PreMem*	37.4	27.3	26.5	43.8	**38.1**

K Failure Mode Analysis

Beyond the comparison in Tab. 6, manual inspection of DualMem-RL outputs surfaces three recurring failure modes: two in memory *utilization* (F1, F2) and one in memory *construction* (F3).

F1: Retrieved Insight Not Recognized as Relevant

- *Persona:* disability vocational counselor. Core value: professional confidence stems from self-identity, not external pity.
- *Insight:* organizations default coordination and team-building roles to women based on structural bias rather than ability assessment.

- *Query:* "if my leader assigns team-building again, should I refuse?"
- *Response:* generic both-sides advice ("it depends; team building improves communication; workplace relationships matter").
- *Diagnosis:* the insight reframes the query from a scheduling conflict into a career-equity issue. The agent retreats to AI-style neutrality and leaves the core value entirely unutilized.

F2: Insight Adopted but Supporting Facts Ignored

- *Persona:* animal welfare coordinator. Core value: sustainable support networks require rescuers' own psychological resilience.
- *Insight:* the one-directional giving pattern systematically depletes psychological resilience.
- *Supporting facts:* the station leader funds all supplies and medical costs out of pocket without seeking external help; a young volunteer with a wrist injury kept working to avoid burdening others, then quietly withdrew.
- *Query:* "How should I design a psychological support activity so that rescuers are willing to attend?"
- *Response:* set up a relaxation corner at adoption events with meditation guidance and photo sharing.
- *Diagnosis:* the response implicitly adopts the insight's direction but cites none of the grounding facts. Without the concrete evidence of self-funded exhaustion and injury-driven withdrawal, the recommendation addresses participation logistics rather than the deeper barrier of rescuers' internalized obligation to give without receiving.

F3: Surface Similarity Conflation in Insight Construction

- *Persona:* ancient manuscript restorer. Core value: minimal intervention, restoring artifacts without subjective alteration.
- *Facts:* the persona *passively noticed* deterioration in a late-Qing account book found at an old residence; separately, when selecting koi, the persona *actively handled* fish before purchase as hands-on verification.
- *Constructed insight:* the persona acquires first-hand information through direct sensory contact, reflecting a hands-on practical spirit.
- *Diagnosis:* the two facts differ fundamentally in behavioral nature, passive discovery versus deliberate active verification. The model conflates them on the surface similarity of physical contact and promotes an unwarranted stable trait, misrepresenting the persona's decision process and risking overgeneralized responses when retrieved.

L Human Evaluation Details

For the human evaluation phase, we recruited two annotators with prior experience in NLP annotation tasks. Both annotators are native speakers based in mainland China. To ensure ethical research practices and high-quality data, we provided compensation at a rate of 100 RMB (approximately $14 USD) per hour. This rate is higher than the local minimum wage and is commensurate with the expert nature of the task. Furthermore, we adhered to privacy protocols: no personally identifiable information was collected, and all evaluation data were anonymized.