PERSONAMEM-V2: Towards Personalized Intelligence via Learning Implicit User Personas and Agentic Memory

Anonymous ACL submission

Abstract

Personalization is one of the next milestones in advancing AI capability and alignment. We introduce PERSONAMEM-V2, the state-of-the-art dataset for LLM personalization that simulates 1,000 realistic user–chatbot interactions on 300+ scenarios, 20,000+ user preferences, and 128k-token context windows, where most user preferences are implicitly revealed to reflect real-world interactions. Using this data, we investigate how reinforcement fine-tuning enables a model to improve its long-context reasoning capabilities for user understanding and personalization. We also develop a framework for training an agentic memory system, which maintains a single, human-readable memory that grows with each user over time. In our experiments, frontier LLMs still struggle with implicit personalization, achieving only 37–48% accuracy. Using reinforcement fine-tuning, we successfully train Qwen3-4B to outperforms GPT-5, reaching 53% accuracy in implicit personalization. Moreover, our agentic memory framework achieves state-of-the-art 55% accuracy while using 16× fewer input tokens, relying on a 2k-token memory instead of full 32k conversation histories. These results underscore the impact of our dataset and demonstrate agentic memory as a scalable path toward real-world personalized intelligence.

1 Introduction

Personalization is becoming one of the next milestones towards artificial super-intelligence (OpenAI, 2025,b; Meta, 2025; Chowdhury, 2025; Evolution AI Hub, 2025). As the user base of artificial intelligence grows dramatically in both size and diversity, large language models (LLMs) increasingly face the challenge of serving users with distinct personas and backgrounds. In many real-world applications, such as education, healthcare, and empathetic emotional support (Jiang et al., 2025; Ghimire et al., 2024; Ivanovic et al., 2022; Gómez-González et al., 2020; Baillifard et al., 2025; Schaaff et al., 2023), there is no single "correct" answer. Instead, success depends on delivering personalized responses that align with individual users' intentions, contexts, preferences, and emotional states. **Personalization offers a path toward pluralistic alignment**, shifting the reasoning goals towards factual correctness to meaningful resonance across diverse users.

A key enabler of this vision is the growing history of user–chatbot interactions, which encode rich signals about user preferences. However, personalization in LLMs remains a challenge. Most users do not explicitly state their preferences to chatbots. According to OpenAI's recent report (OpenAI, 2025), most users still treat LLMs as tools, meaning most of their preferences are revealed only implicitly through everyday interactions. For example, someone might ask a chatbot to help polish the writing of an email, but the email itself could reveal their dining habits, as shown in Figure 1. As a result, real-world conversation histories tend to be long and noisy, requiring models to infer implicit user personas and preferences from scattered, indirect evidence over time.

This raises a central question: **how well can LLMs understand the users, especially their implicit personas and preferences from long conversation histories, and therefore provide personalized responses?**

To this end, we present the state-of-the-art LLM personalization dataset, PERSONAMEM-V2: IMPLICIT PERSONAS, that captures the complexity of real-world user–chatbot interactions. Our dataset spans over 1,000 user personas, covering comprehensive demographic attributes, mental health and medical backgrounds, as well as stereotypical, anti-stereotypical, and neutral user preferences. In addition, it features realistic, dynamic, multimodal, multilingual, and multi-session user–chatbot conversation histories that implicitly convey user pref-

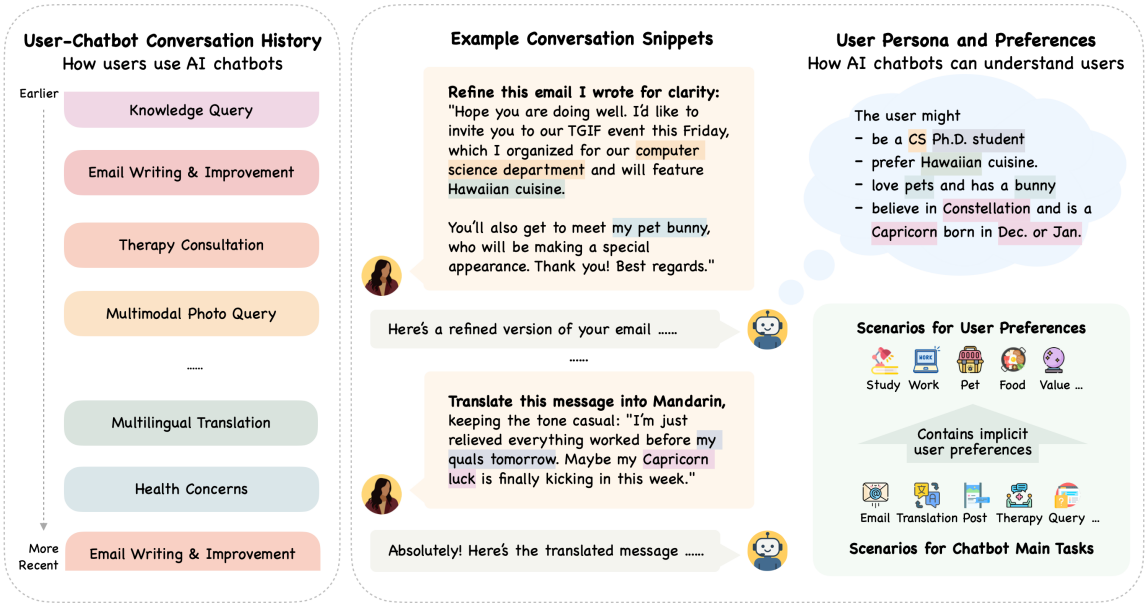

Figure 1: Overview of PERSONAMEM-V2. On the left, we mimic how people talk to chatbots across many topics over time, forming long and varied interaction histories. In the middle, we zoom in on a few conversation snippets. Even simple task scenarios like refining an email naturally reveal small details about someone's life in many other scenarios, like what they study, what foods they like, whether they have pets, or what they're planning for tomorrow. Taken together, these subtle signals help models build thorough yet succinct user profiles for personalization.

erences, encompassing over 20,000+ user preferences over 300+ everyday conversation topics, and up to 128,000 tokens per context. To support both evaluation and model training, the dataset includes 5,000 high-quality Q&A pairs curated for benchmarking and an additional 20,000 Q&A pairs for training and validation, each generated through a scalable, multi-step validation pipeline that enforces strong quality standards. We show all related work in Appendix A.1.

To understand how well AI models personalize to users, we started by testing frontier LLMs from OpenAI (OpenAI, 2025a, 2024b,a) on these implicit signals that show up in everyday conversations. These evaluations quickly revealed a gap: even the strongest models among them struggle to interpret and track user preferences over long interaction histories. This gap motivated us to explore the value of our data in driving the next wave of personalized intelligence. Using reinforcement fine-tuning (RFT) with GRPO (Guo et al., 2025), we find that RFT is remarkably effective for personalization, even though personalization is inherently subjective and open-ended. Our dataset enables us to train a 4B-parameter reasoning model capable of long-context reasoning for personalization.

Another key contribution of our work is an agentic memory framework that learns to build and update a human-readable memory that evolve with each user. Instead of relying on full conversational transcripts, the model is trained via RFT to distill long histories into a compact 2k-token memory. This memory becomes the model's sole personalization context, allowing it to reason over implicit user personas and preferences and maintain strong user understanding across multi-session conversations. Despite using 16× fewer input tokens, the agentic memory model delivers state-of-the-art performance on implicit personalization. This demonstrates a scalable path toward AI systems that remember what matters for each user, infer subtle user preferences, and adapt over time to give personalized responses, all with the efficiency required for real-world deployment. To summarize our contributions:

- We curate the state-of-the-art dataset for LLM personalization, featuring realistic user–chatbot interactions that reveal implicit user preferences.
- We benchmark the implicit personalization capabilities of frontier LLMs.
- We demonstrate the efficacy of reinforcement fine-tuning for personalization.
- We propose an agentic memory framework that achieves best performance and efficiency.

2 Overview of PERSONAMEM-V2: IMPLICIT PERSONAS

2.1 Comprehensive User Personas

PERSONAMEM-V2 introduces **1,000 richly detailed personas** that capture nearly the full spectrum of demographic diversity across global regions, cultures, races, genders, and sexual orientations, reflecting the pluralism that real-world AI systems must serve. Each persona draws from a random description in PERSONAHUB (Ge et al., 2024) and expands it into an unfixed set of attributes that cover, but are not limited to, professional and educational backgrounds, personal characteristics, relationships, values and beliefs, technological familiarity, and conversational styles. Moreover, as people increasingly rely on chatbots for personal use and care, PERSONAMEM-V2 specifically includes mental and physical health backgrounds where personalization carries high stakes.

A central question for LLM personalization is what a chatbot should remember about the user. The chatbot should remember what the user has shared about themselves, whether conveyed implicitly or explicitly, rather than relying on static assumptions or stereotypes derived solely from demographic attributes. To explore this boundary, PERSONAMEM-V2 provides **stereotypical, anti-stereotypical, and neutral user preferences** for each persona, as well as health- and therapy-related ones, totaling around 20,000+, enabling systematic evaluation of how models ground personalization in conversational evidence.

Chatbots often misinterpret user behavior. This can occur when people test the model with hypothetical examples to explore its capabilities, or with messages written by a third person, and the model fails to distinguish them from the user's own persona. PERSONAMEM-V2 addresses this challenge by including such ambiguous cases. Besides, we also account for the **dynamic nature of user preferences and how they evolve over time** across multiple sessions. For instance, a user who initially expresses interest in vegetarian recipes might later ask for high-protein meal suggestions after starting a new fitness routine.

2.2 Multi-Session Conversation Histories

Each user preference is converted into a multi-turn conversation ranging from two to six turns, simulating how users naturally interact with chatbots over time. A key design principle is **cross-scenario personalization**: users rarely state who they are or what they like directly to a chatbot. Instead, they treat chatbots as tools for everyday tasks, such as writing emails, translating text, or seeking information (OpenAI, 2025). Yet within these tasks, subtle cues about their personas and preferences often emerge, implicitly revealing aspects of their background, hobbies, or cultural values (Liu et al., 2025c), for example, their field of study, favorite cuisine, or even the belief in astrology. As illustrated in Figure 1, this cross-scenario dynamic allows the chatbot to learn user preferences not from direct self-descriptions, but from natural, task-driven interactions. By modeling such patterns, PERSONAMEM-V2 captures how personalized signals arise organically across multiple contexts, enabling reasoning that generalizes beyond a single use case.

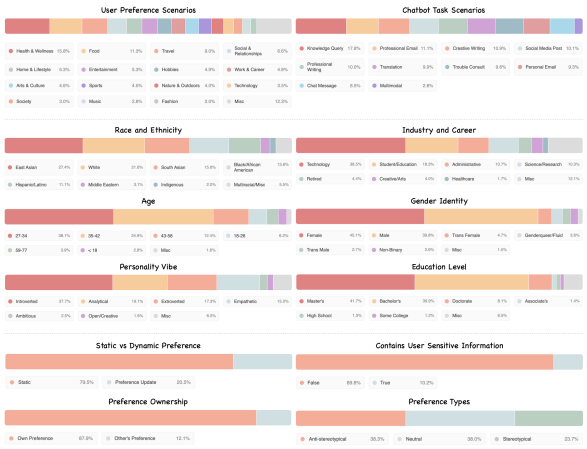

Figure 2: What's inside the PERSONAMEM-V2 dataset. It spans broad and diverse distributions of user personas, preferences, and task scenarios, ranging from everyday interests like food, travel, and hobbies to demographic attributes, personality vibes, and professional backgrounds. This wide coverage is designed to better support training and evaluating personalized AI systems that reflect real-world users and use cases.

PERSONAMEM-V2 also offers broad coverage of the everyday tasks people bring to chatbots in daily life, including **writing and improving emails, composing chat messages and social media posts, multilingual translation, multimodal photo query, knowledge exploration, therapy and reflection, and medical consultations**. For instance, a user's photo might hint at their location and activities, and repeatedly asking questions in the same domain can reveal their personal interests or professional focus. Each conversation is also categorized by the detailed topic of its content, resulting in **325 distinct topics** in total.

Our goal is to approximate real-world user–chatbot interactions, which often span multiple sessions over time. To reflect this, we build multi-session histories by concatenating multi-turn conversations from the previous stage, up to 32,000 tokens for each user. Conversation segments are arranged in a topological order, such as preference updates or user requests to forget prior preferences. Additionally, we create a complementary set of dialogues focused on code debugging and mathematical problem-solving tasks. These are independent of user preferences in our dataset and extend the effective context window to **128,000 tokens**, enabling evaluation of long-range reasoning and personalization.

2.3 User Privacy-Aware Design

PERSONAMEM-V2 simulates realistic privacy risks by introducing scenarios where users unintentionally share **sensitive or private information**, such as personal addresses, phone numbers, contact details, or even API keys within conversational contexts. A responsible model is not supposed to leverage such information to generate personalized responses. Besides, users should retain a degree of control over personalization by simply asking the chatbot **not to remember** a particular preference or piece of information. We deliberately include such interactions in our dataset.

2.4 In-Situ User Queries

We simulate how users naturally pose queries to chatbots. Given each ground-truth user preference, with its corresponding conversation snippet somewhere in the history, we generate a Q&A pair, and append the query at the end of the conversation to reflect a current in-situ interaction. We adopt both open-ended and multiple-choice (MCQ) formats, together with a rich set of annotations. We ensure that all four options in MCQ are reasonable, but only one is personalized to the current user. In total, the dataset encompasses 335 user query topics, with 5,000 Q&A pairs in benchmarking, 18,000 in training, and 2,000 in validation to support large-scale training and evaluation.

2.5 Ensuring High Quality of Data

The core principle behind our data pipeline is simple: keep scaling up generation, impose comprehensive quality filtering, and only retain data that pass every filter. We show human evaluation results on data quality in Appendix A.2.

Every Q&A pair undergoes strict validation based on the following principles: (1) The chatbot shall not be able to answer the user's query correctly without seeing the conversation history, avoiding question leaks or artifacts; (2) the correct option must faithfully reflect the user's true preference; (3) no incorrect option may do so; and (4) the formatting must be clean and natural, without artifacts like *"Sure, here is the answer."* Only around 30% of generated Q&As survive all filters, ensuring exceptional quality of the remained data.

The pipeline is designed to scale continuously: we can generate more data, introduce new filters, and tighten evaluation criteria. Each filtering stage aggregates multiple LLM-as-a-judge votes, and all data generation and filtering are conducted with GPT-5 (OpenAI, 2025a) without auto-routing to maintain consistency and top-tier quality.

3 Towards Personalized Intelligence

A main challenge in achieving personalized intelligence is inferring a user's implicit persona and preferences from long, noisy conversational histories. Unlike classical question answering that focuses on retrieving explicit factual information, personalization requires strong reasoning capabilities to extract inherent user preference. Reinforcement learning (RL) has proven effective in enhancing the reasoning abilities of language models across diverse reasoning tasks (Guo et al., 2025; Shen et al., 2025; Zha et al., 2025). Building on this insight, we demonstrate that RL can also drive models toward better personalization.

3.1 RL with Long-Context Reasoning

A natural approach is to treat personalization as a long-context reasoning problem: given an extensive user–chatbot conversation history and a new user query, the model is trained via RL to reason over the full context and generate responses aligned with user preferences.

We adopt Group Relative Proximal Optimization (GRPO) as our reinforcement fine-tuning algorithm (Guo et al., 2025; Shao et al., 2024). To effectively drive personalization, reinforcement fine-tuning requires verifiable rewards that determine whether each model output is aligned with, or contradicts, the user's current persona and preferences. To support this, PERSONAMEM-V2 provides two reward pathways, as illustrated in Figure 3. Each Q&A instance in PERSONAMEM-

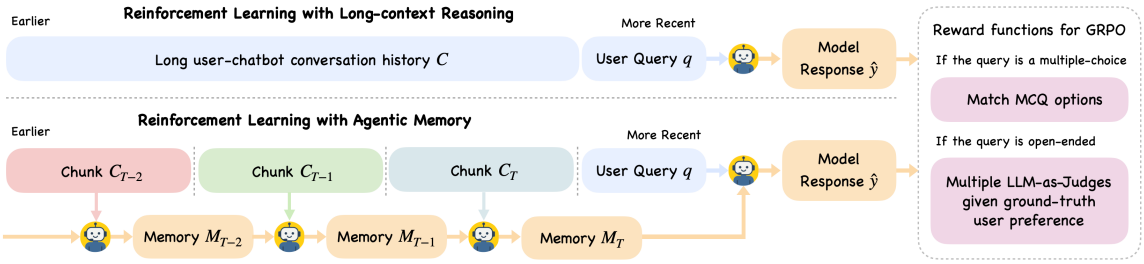

Figure 3: Schematic of our RL-based training strategies. The top figure illustrates long-context reasoning, where the model reasons over the full conversation history. The bottom one depicts agentic memory, where the model divides the full history into chunks and iteratively updates a memory of capped size. The model will receive a high reward if the memory turns out to be necessary and helpful in answering the final user query.

v2 can a multiple-choice task, allowing us to directly measure correctness. It also includes an annotated ground-truth user preference, enabling an LLM-as-a-judge to assess whether an open-ended response reflects that preference.

This straightforward approach naturally encourages the model to reason over long contexts, leveraging the full interaction history to infer subtle, implicit user preferences. However, it also requires appending the entire conversation as context, and as the conversation grows, this quickly becomes inefficient, setting the stage for the need for a more scalable mechanism.

3.2 RL with Agentic Memory: Toward Scalable Personalization

As conversations accumulate, it becomes necessary to maintain a distilled representation of the user's preferences that can be continually updated based on new interactions. Agentic memory (Yu et al., 2025; Xu et al., 2025; Zhong et al., 2024; Zhang et al., 2025d; Li et al., 2025b) offers a scalable mechanism for language models to distill long-term user information into a compact user persona. Unlike long-context QA (Yang et al., 2018), which focuses on retrieving explicit facts, personalization requires reasoning about subtle, implicit preference signals to write a memory, and track how these preferences evolve over time to update it, making personalization a continuous reasoning problem that agentic memory is naturally equipped to address.

Inspired by MEMAGENT (Yu et al., 2025), we divide the entire conversation history into fixed-size chunks. The objective is to sequentially construct and update a single agentic memory M from a growing set of T chunks. In practice, each chunk may correspond to several days or weeks of interactions. To support personalization, we follow three core principles motivated by real-world settings:

- **Causality** – the model should only write memory based on what the user has said so far, never on future chunks or user queries it has not yet seen.
- **Markovian assumption** – the memory at step i must summarize everything that matters from the past, so the next update only depends on C_i and M_{i-1}.
- **Capped memory size** – the memory must stay compact and human-readable, enabling the system to scale to long-term use without growing unbounded.

These constraints force the model to decide what to store without knowing future user queries, encouraging it to anticipate what information will matter and to extract the essential user persona. At each step, the model reads the current chunk C_i and the previous memory M_{i-1} to produce an updated memory. After processing all chunks, the same model answers the final query q using the memory M_T. The pipeline is shown in Figure 3. Formally:

$$M_i = f_\theta(C_i,\, M_{i-1}), \quad \hat{y} = f_\theta(M_T,\, q)$$

where f_θ is the language model, C_i is the i-th chunk out of T, M_{i-1} and M_i are the previous and updated memories, q is the current user query, and $\hat{y}$ is the model response to q.

We use a single LLM for the entire process: the same model that writes the memory also gives the final answer. All queries in PERSONAMEM-V2 can only be answered in a personalized manner by having a correct user understanding, so the model will receive high rewards if the memory turns out to be both necessary and helpful in giving correct personalized responses, and vice versa. **Over time, the model learns to reason over context and maintain a concise, human-readable memory that grows with each user**.

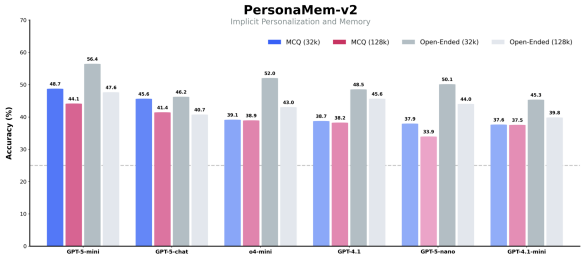

Figure 4: Performance of OpenAI models on the PERSONAMEM-V2 benchmark, comparing accuracy across 32k and 128k input contexts in MCQ and Open-Ended tasks. Despite recent advancements, we observe that frontier models still struggle with implicit personalization. The dashed line represents the random guess performance in MCQ tasks.

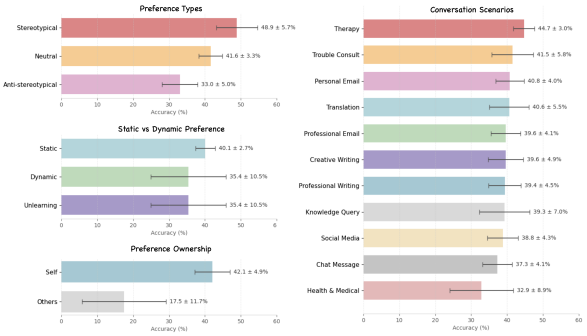

Figure 5: Breakdown of model accuracy by preference types and scenarios, aggregated across all evaluated models. The results indicate systematic variance in personalization capabilities, showing a reliance on population priors, with performance dropping for anti-stereotypical and dynamic preferences and distinguishing preference ownerships.

4 Experimental Results

4.1 Benchmarking Frontier LLMs in Personalization

We benchmark a series of OpenAI models, including GPT-5-Chat, GPT-5-mini, GPT-5-nano, GPT-4.1, GPT-4.1-mini, and o4-mini (OpenAI, 2025a, 2024b). Each model is evaluated under both multiple-choice (MCQ) and open-ended settings. For open-ended evaluations, we adopt an LLM-as-a-judge protocol: three independent GPT-5-Chat instances score each response, and we aggregate their judgments to obtain more robust accuracy estimates. We show overall results in Figure 4 and error analysis in Appendix A.3.

4.1.1 Frontier LLMs still struggle to infer implicit user preferences

Despite major advances in long-context handling, frontier LLMs still perform poorly when required to infer implicit user preferences. Across both MCQ and open-ended settings, which show strong correlation, GPT-5 variants reach only 40-55% accuracy. This reveals an obvious gap in current frontier models' ability to understand user preferences and offer personalized responses based on subtle cues in interaction history.

4.1.2 Reasoning, not long-context capabilities, drives success in implicit personalization

The models that lead the benchmark like GPT-5-mini and o4-mini are those with strong reasoning capabilities. Meanwhile, perhaps surprisingly, we observe no significant accuracy improvement when the context is shortened from 128k to 32k tokens by removing conversations irrelevant to the current persona. These results suggest that the main bottleneck in implicit personalization is not simply memorizing longer histories or retrieving factual information, but correctly interpreting and integrating subtle preference signals embedded within those histories. We need to update our focus: moving from "needle-in-a-haystack" (Kamradt, 2023; Team et al., 2024) retrieval-style stress tests toward more nuanced, fine-grained assessments that reflect the demands of real-world personalization.

4.1.3 Implicit personalization varies systematically across preference types

Figure 5 shows that implicit personalization varies systematically across preference types. Frontier models perform best when user preferences align with stereotypes, achieving 48.9 ± 5.7% accuracy aggregated across all models and context window lengths, but it drops to 41.6 ± 3.3% for neutral preferences and further to 33.0 ± 5.0% for anti-stereotypical preferences, suggesting reliance on population priors rather than individual behavior.

Besides, static preferences reach 40.1 ± 2.7% accuracy, while dynamic preferences fall to 35.4 ± 10.5%, indicating instability and difficulty updating beliefs as preferences change. For preference ownership, models infer users' own preferences correctly at 42.1 ± 4.9%, but perform far worse, only 17.5 ± 11.7%, when distinguishing preferences of others from the user's own. Differences across task scenarios are comparatively modest, ranging from 44.7 ± 3.0% in therapy consultation to 32.9 ± 8.9% in physical health and medical scenarios.

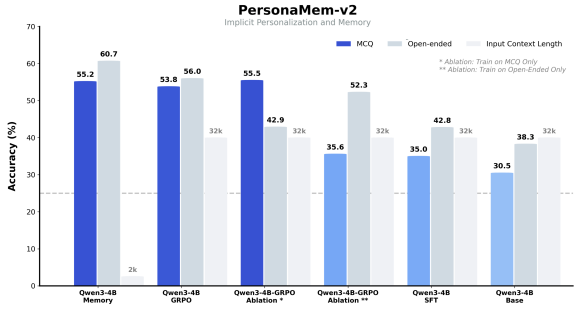

Figure 6: Performance of our Qwen3-4B models on PERSONAMEM-V2 trained via SFT, GRPO, and Agentic Memory, alongside ablation studies for GRPO trained solely on MCQ or Open-Ended data. Comparisons show that RL drives significant gains in implicit personalization. Notably, our Agentic Memory framework achieves both SOTA performance and efficiency using only 2k input tokens as memory throughout the 32k token history. The dashed line represents the random guess performance in MCQ tasks.

4.2 Training Long-context Reasoning for Personalization

We begin with Qwen3-4B-Instruct-2507 (Yang et al., 2025) and train it with verl: Volcano Engine Reinforcement Learning for LLMs (Sheng et al., 2025) and vLLM (Kwon et al., 2023) on 8 Nvidia H100 GPUs. The training dataset contains 18,000 samples with no persona overlap with the benchmark set. All models are trained on this dataset and evaluated on the benchmark set, using a context window of up to 32k tokens. The model is first cold-started with 300 steps of supervised fine-tuning with a batch size of 64 on our training data, then reinforced with GRPO for 1 epoch around 500 steps, enabling thinking and using a batch size of 32 with 8 rollouts. We mix 80% multiple-choice and 20% open-ended queries to keep training stable with more verifiable rewards from MCQ while still preserving open-ended conversational abilities for real-world personalization. We use three GPT-5-Chat instances as LLM-as-judges to provide rewards on the open-ended samples.

4.2.1 Reinforcement learning incentives reasoning toward personalization

While personalization is often viewed as subjective and pluralistic, we find that it can still be effectively incentivized through reinforcement fine-tuning. In particular, well-designed reward signals encourage models to reason about whether a response aligns with or violates the current user's persona and preferences. Our dataset plays a crucial role here: it provides the structured yet diverse supervision needed for RL to learn these behaviors. Using this data, our 4B-parameter reasoning model, marked as Qwen3-4B-GRPO in Figure 6, gains substantial improvements after post-training, reaching 53.8% on MCQ and 56.0% on open-ended tasks, superior to supervised fine-tuning. Notably, this small model even outperforms GPT-5-Chat, which achieves 45.6% and 46.2% on the same benchmarks. These results show that reinforcement learning can meaningfully shape a model's ability to interpret and integrate subtle user preferences to give personalized responses.

4.2.2 Hybrid reward signals unlock better RL toward personalization

To understand how different supervision signals shape reinforcement fine-tuning for personalization, we ran ablations comparing the same Qwen3-4B models trained only on MCQ, only on open-ended questions, or on our mixed dataset, keeping the same training setup, GRPO algorithm, and total number of training samples. The results shown in Figure 6 are striking: the MCQ-only model collapsed by 13.1% on open-ended personalization tasks needed for real-world user-facing services, while the open-ended–only model dropped 18.2% on MCQs and even 3.7% on open-ended evaluation itself. We hypothesize that MCQs provide the stable, verifiable rewards that reinforcement fine-tuning depends on, whereas open-ended samples with LLM-as-a-judge capture the richness needed for teaching nuanced conversational behavior but can make reward signals less stable. Mixing both types of questions offers the best of both worlds, providing reliable reward anchors while preserving the flexibility required for nuanced, real-world user understanding and personalization.

4.3 Training Agentic Memory for Personalization

We follow the multi-turn conversation RL training framework of MEMAGENT (Yu et al., 2025). Each iteration of the memory-update process can be viewed as an independent conversation: the model receives a chunk, updates the memory, and the optimization is performed separately for each conversation. The RL objective follows GRPO (Guo et al., 2025), where the advantage of each rollout is computed using the reward received at the final turn, which will be shared across all preceding conversations. We show related prompts in Appendix A.4.

..... You appreciate thoughtful discussion and have shown a balance between engaging in academic pursuits and pursuing **personal interests such as cooking**, which you see as a way to reflect on patience and harmony. ..... You have expressed a preference for **outdoor activities in the spring**, such as walking by a river, visiting a botanical garden's indoor greenhouse, or **practicing photography** in shaded urban parks, while avoiding grassy or flower-heavy areas due to a mild **seasonal pollen allergy**. .....

Figure 7: Inside the memory: a glimpse of the fine-grained user personas learned from conversation histories through our RL-based agentic memory framework.

To maintain consistency with the long-context reasoning setup, we train on the same training framework and the mixed dataset consisting of 80% multiple-choice and 20% open-ended queries. As before, we initialize from Qwen3-4B-Instruct-2507 after a cold-start supervised finetuning stage. For RL training, we set the batch size to 32 with 8 rollouts per prompt, and train for approximately 500 steps. Under the 32k-token context window, we cap the memory size M_i at 2,048 tokens to ensure compactness and efficiency, while each chunk C_i is limited to 5,000 tokens, yielding a total of $T = 8$ memory-update iterations.

4.3.1 Agentic memory delivers state-of-the-art performance with unmatched efficiency

Our agentic memory model achieves the strongest personalization performance across all evaluated models. As shown in Figure 6, it reaches 55.2% accuracy on MCQ and 60.7% on open-ended evaluations, surpassing both the same Qwen3-4B model trained directly with long-context reasoning and the GPT-5 series of frontier models.

Crucially, this performance comes with dramatic efficiency gains. Instead of repeatedly processing full 32k-token conversation histories, the model relies on a compact 2k-token memory throughout the process, making it 16× more efficient without even sacrificing the performance. This makes agentic memory well-suited for real-world personalized AI deployments where latency, cost, and context limits are major constraints.

4.3.2 Human-readable memory enables transparency and user control

Beyond raw performance, agentic memory introduces a new dimension to personalization: transparency and user control. Specifically, our memory framework maintains a human-readable memory that evolves with each user over time, allowing users to audit the memory, correct misunderstandings, and guide the model's personalization behavior directly. We show a glimpse of memory in Figure 7. This explicit memory format also opens the door to new deployment strategies. We hypothesize a practical scenario where AI systems can maintain and update user memory offline, similar to a sleep-time compute cycle (Lin et al., 2025), providing a smooth user experience and practical personalized intelligence.

5 Conclusion and Future Work

We introduce PERSONAMEM-V2: IMPLICIT PERSONAS, the state-of-the-art dataset designed for implicit personalization over long context, capturing 1,000 user personas and their realistic, dynamic, multimodal, multilingual, and multi-session user–chatbot interactions across 300+ conversation scenario. These conversations span 20,000+ cross-scenario user preferences from utility task–driven interactions like writing improvement and translation, reflecting how real-world users engage with AI systems in their everyday lives.

Our results show that current frontier LLMs still struggle to interpret implicit user preferences. Besides, reasoning, not longer-context handling or memorization in frontier models, drives success in personalization. Existing personalization also varies systematically across preference types, suggesting reliance on population priors rather than individual behavior. With targeted reinforcement fine-tuning, a 4B reasoning model can outperform GPT-5, and our agentic memory framework further pushes performance to the state-of-the-art while remaining dramatically 16x more efficient, providing a scalable path toward real-world personalized intelligence.

Looking ahead, we see exciting opportunities: richer multimodal user personalization, more structured and interactive memory architectures, personalization over more utility tasks, user-customizable boundary between personalization and privacy, and leveraging real user–chatbot interactions to build even more realistic training data. Overall, our findings point toward a future of personalized intelligence and agentic memory framework that can remember, reason, and adapt to individual users over long contexts, enabling more pluralistic alignment and deeper personalized resonance in future intelligent systems.

6 Limitations

6.1 Synthetic Personas and Conversations

PERSONAMEM-V2 is constructed using synthetic personas and conversations powered by GPT-5, which enables scalable, controllable, and reproducible evaluation of personalization. While synthetic data may not fully reflect all aspects of real user behavior, it provides a practical and widely adopted alternative when large-scale real-user data is difficult and expensive to obtain. Collecting real-world personalized interaction data can also raise user privacy concerns, and require substantial effort in data cleaning due to the noise and inconsistency in real data. Synthetic data thus offers a complementary approach that allows controlled analysis while avoiding these challenges.

6.2 Coverage of Personalization Scenarios

The benchmark covers a broad range of personalization tasks and preference types, aiming to reflect common and representative usage patterns, such as email refinement, translation, knowledge queries, photo sharing, and therapy-style consultation. Nevertheless, real-world personalization can involve more complex, evolving, and highly domain-specific behaviors that are difficult to capture comprehensively in a single benchmark. It is not designed to directly evaluate highly specialized or safety-critical domains, e.g., clinical or legal decision-making, nor agentic settings that require extensive tool use, long-horizon planning, or long-form document generation. These settings represent important but orthogonal directions that we leave to future benchmarks.

6.3 Agentic Memory Design Choices

Our work investigates the potential of agentic memory mechanisms for long-context personalization, but does not aim to exhaustively explore the full memory design space. Alternative approaches, such as hierarchical memories, more structured memory representations, separate memories for short-term and long-term information, or RAG-based memory, may further enhance performance and are promising directions for future research. Importantly, the proposed framework is intentionally flexible: different memory representations can be achieved through different prompting strategies and format-based rewards, allowing practitioners to customize memory behavior to different application needs.

7 Ethical Considerations

7.1 Personalization and User Privacy

Personalized intelligence must balance usefulness with user privacy. In this work, we explicitly incorporate privacy considerations into the benchmark design and evaluation. First, PERSONAMEM-V2 does not rely on real user data; all personas and conversation histories are fully synthesized, eliminating risks associated with collecting or exposing real personal information. Second, we include pseudo sensitive information, e.g., home addresses, phone numbers, identification numbers, API keys, and so on to mimic realistic scenarios in which such information may appear in user–assistant interactions. The target responses used to train the models are constructed to intentionally avoid utilizing on this information to give personalized responses. Third, we include scenarios in which users explicitly request the assistant to forget or refrain from using specific preferences, encouraging models to respect user intent and privacy controls. Finally, the proposed agentic memory module stores user information in a transparent and human-readable way, allowing users to inspect, monitor, and edit the recorded memory. Together, these design choices aim to support personalization research while encouraging user privacy-aware designs.

7.2 Residual Biases in Synthetic Data

While we aim to cover diverse personas and preferences, synthetic data generation may still introduce residual biases or stereotypes. To mitigate this to our best, we have apply a series of GPT-5–powered filtering and validation steps throughout the data generation process to remove harmful and low-quality samples. In addition, when generating both stereotypical and anti-stereotypical attributes for each persona, we explicitly ensure that all resulting content remains non-harmful and appropriate. Addressing more subtle or emergent biases in user persona and preference distributions remains an important direction for future dataset refinement.

References

Ambroise Baillifard, Maxime Gabella, Pamela Banta Lavenex, and Corinna S Martarelli. 2025. Effective learning with a personal ai tutor: A case study. *Education and Information Technologies*, 30(1):297–312.

Hyungjune Bu, Chanjoo Jung, Minjae Kang, and Jaehyung Kim. 2025. Personalized llm decoding via contrasting personal preference. *arXiv preprint arXiv:2506.12109*.

Hongru Cai, Yongqi Li, Wenjie Wang, Fengbin Zhu, Xiaoyu Shen, Wenjie Li, and Tat-Seng Chua. 2025. Large language models empowered personalized web agents. In *Proceedings of the ACM on Web Conference 2025*, pages 198–215.

Jiarui Chen. 2025. Memory assisted llm for personalized recommendation system. *arXiv preprint arXiv:2505.03824*.

Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav. 2025. Mem0: Building production-ready ai agents with scalable long-term memory. *arXiv preprint arXiv:2504.19413*.

Shubhangi Chowdhury. 2025. Gpt-6 preview: Openai's big step toward personalized ai. https://americanbazaaronline.com/2025/08/20/gpt-6-preview-openais-big-step-toward-personalized-ai-466456/. Accessed: 2025-10-20.

Evolution AI Hub. 2025. Gpt-6 leak: Openai's next ai can remember everything about you. *Medium*. Accessed: 2025-10-20.

Tao Ge, Xin Chan, Xiaoyang Wang, Dian Yu, Haitao Mi, and Dong Yu. 2024. Scaling synthetic data creation with 1,000,000,000 personas. *arXiv preprint arXiv:2406.20094*.

Aashish Ghimire, James Pather, and John Edwards. 2024. Generative ai in education: A study of educators' awareness, sentiments, and influencing factors. In *2024 IEEE Frontiers in Education Conference (FIE)*, pages 1–9. IEEE.

Emilio Gómez-González, Emilia Gomez, Javier Márquez-Rivas, Manuel Guerrero-Claro, Isabel Fernández-Lizaranzu, María Isabel Relimpio-López, Manuel E Dorado, María José Mayorga-Buiza, Guillermo Izquierdo-Ayuso, and Luis Capitán-Morales. 2020. Artificial intelligence in medicine and healthcare: a review and classification of current and near-future applications and their ethical and social impact. *arXiv preprint arXiv:2001.09778*.

Jian Guan, Junfei Wu, Jia-Nan Li, Chuanqi Cheng, and Wei Wu. 2025. A survey on personalized alignment–the missing piece for large language models in real-world applications. *arXiv preprint arXiv:2503.17003*.

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: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Jiani Huang, Xingchen Zou, Lianghao Xia, and Qing Li. 2025. Mr. rec: Synergizing memory and reasoning for personalized recommendation assistant with llms. *arXiv preprint arXiv:2510.14629*.

Mirjana Ivanovic, Serge Autexier, and Miltiadis Kokkonidis. 2022. Ai approaches in processing and using data in personalized medicine. In *European Conference on Advances in Databases and Information Systems*, pages 11–24. Springer.

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

Gregory Kamradt. 2023. Needle in a haystack - pressure testing llms. https://github.com/gkamradt/LLMTest_NeedleInAHaystack.

Saeed Khaki, JinJin Li, Lan Ma, Liu Yang, and Prathap Ramachandra. 2024. Rs-dpo: A hybrid rejection sampling and direct preference optimization method for alignment of large language models. *arXiv preprint arXiv:2402.10038*.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th symposium on operating systems principles*, pages 611–626.

Hao Li, Chenghao Yang, An Zhang, Yang Deng, Xiang Wang, and Tat-Seng Chua. 2025a. Hello again! llm-powered personalized agent for long-term dialogue. 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 5259–5276.

Xinyu Li, Ruiyang Zhou, Zachary C Lipton, and Liu Leqi. 2024. Personalized language modeling from personalized human feedback. *arXiv preprint arXiv:2402.05133*.

Zhiyu Li, Shichao Song, Hanyu Wang, Simin Niu, Ding Chen, Jiawei Yang, Chenyang Xi, Huayi Lai, Jihao Zhao, Yezhaohui Wang, and 1 others. 2025b. Memos: An operating system for memory-augmented generation (mag) in large language models. *arXiv preprint arXiv:2505.22101*.

Kevin Lin, Charlie Snell, Yu Wang, Charles Packer, Sarah Wooders, Ion Stoica, and Joseph E Gonzalez. 2025. Sleep-time compute: Beyond inference scaling at test-time. *arXiv preprint arXiv:2504.13171*.

Jiahong Liu, Zexuan Qiu, Zhongyang Li, Quanyu Dai, Wenhao Yu, Jieming Zhu, Minda Hu, Menglin Yang, Tat-Seng Chua, and Irwin King. 2025a. A survey of personalized large language models: Progress and future directions. *arXiv preprint arXiv:2502.11528*.

Zijun Liu, Peiyi Wang, Runxin Xu, Shirong Ma, Chong Ruan, Peng Li, Yang Liu, and Yu Wu. 2025b. Inference-time scaling for generalist reward modeling. *arXiv preprint arXiv:2504.02495*.

Ziyi Liu, Priyanka Dey, Zhenyu Zhao, Jen-tse Huang, Rahul Gupta, Yang Liu, and Jieyu Zhao. 2025c. Can llms grasp implicit cultural values? benchmarking llms' metacognitive cultural intelligence with cq-bench. *arXiv preprint arXiv:2504.01127*.

Adyasha Maharana, Dong-Ho Lee, Sergey Tulyakov, Mohit Bansal, Francesco Barbieri, and Yuwei Fang. 2024. Evaluating very long-term conversational memory of llm agents. *arXiv preprint arXiv:2402.17753*.

Reza Yousefi Maragheh, Pratheek Vadla, Priyank Gupta, Kai Zhao, Aysenur Inan, Kehui Yao, Jianpeng Xu, Praveen Kanumala, Jason Cho, and Sushant Kumar. 2025. Arag: Agentic retrieval augmented generation for personalized recommendation. *arXiv preprint arXiv:2506.21931*.

Meta. 2025. Personal superintelligence. https://www.meta.com/superintelligence/?srsltid=AfmBOoqaxZKpn6CgLnzUTxt5HbaHR5jEGRZpPtnpgQZiriZiE78cyUxE. Accessed: 2025-10-20.

OpenAI. 2024a. Gpt-4.1. https://openai.com/index/gpt-4-1/. Accessed: 2025-11-29.

OpenAI. 2024b. Openai o3 and o4-mini system card. https://openai.com/index/o3-o4-mini-system-card/. Accessed: 2025-10-29.

OpenAI. 2025a. Gpt-5 system card. Technical report, OpenAI. Accessed: 2025-10-20.

OpenAI. 2025b. Gpt-5.1. https://openai.com/index/gpt-5-1/. Accessed: 2025-11-15.

OpenAI. 2025. How people are using chatgpt. Accessed: 2025-10-20.

OpenAI. 2025. The power of personalized ai. https://openai.com/global-affairs/the-power-of-personalized-ai/. Accessed: 2025-10-20.

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.

Charles Packer, Vivian Fang, Shishir_G Patil, Kevin Lin, Sarah Wooders, and Joseph_E Gonzalez. 2023. Memgpt: Towards llms as operating systems.

Deeksha Prahlad, Chanhee Lee, Dongha Kim, and Hokeun Kim. 2025. Personalizing large language models using retrieval augmented generation and knowledge graph. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 1259–1263.

Cheng Qian, Zuxin Liu, Akshara Prabhakar, Zhiwei Liu, Jianguo Zhang, Haolin Chen, Heng Ji, Weiran Yao, Shelby Heinecke, Silvio Savarese, and 1 others. 2025. Userbench: An interactive gym environment for user-centric agents. *arXiv preprint arXiv:2507.22034*.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741.

Vishal Raman, Abhijith Ragav, and 1 others. 2025. Remi: A novel causal schema memory architecture for personalized lifestyle recommendation agents. *arXiv preprint arXiv:2509.06269*.

Alireza Salemi, Sheshera Mysore, Michael Bendersky, and Hamed Zamani. 2024. Lamp: When large language models meet personalization. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7370–7392.

Kristina Schaaff, Caroline Reinig, and Tim Schlippe. 2023. Exploring chatgpt's empathic abilities. In *2023 11th international conference on affective computing and intelligent interaction (ACII)*, pages 1–8. IEEE.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, and 1 others. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Maohao Shen, Guangtao Zeng, Zhenting Qi, Zhang-Wei Hong, Zhenfang Chen, Wei Lu, Gregory Wornell, Subhro Das, David Cox, and Chuang Gan. 2025. Satori: Reinforcement learning with chain-of-action-thought enhances llm reasoning via autoregressive search. *arXiv preprint arXiv:2502.02508*.

Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. 2025. Hybridflow: A flexible and efficient rlhf framework. In *Proceedings of the Twentieth European Conference on Computer Systems*, pages 1279–1297.

Taiwei Shi, Zhuoer Wang, Longqi Yang, Ying-Chun Lin, Zexue He, Mengting Wan, Pei Zhou, Sujay Jauhar, Sihao Chen, Shan Xia, and 1 others. 2024. Wildfeedback: Aligning llms with in-situ user interactions and feedback. *arXiv preprint arXiv:2408.15549*.

Fahim Tajwar, Anikait Singh, Archit Sharma, Rafael Rafailov, Jeff Schneider, Tengyang Xie, Stefano Ermon, Chelsea Finn, and Aviral Kumar. 2024. Preference fine-tuning of llms should leverage suboptimal, on-policy data. *arXiv preprint arXiv:2404.14367*.

Gemini Team, Petko Georgiev, Ving Ian Lei, Ryan Burnell, Libin Bai, Anmol Gulati, Garrett Tanzer, Damien Vincent, Zhufeng Pan, Shibo Wang, and 1 others. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*.

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

Zheng Wang, Zhongyang Li, Zeren Jiang, Dandan Tu, and Wei Shi. 2024. Crafting personalized agents through retrieval-augmented generation on editable memory graphs. *arXiv preprint arXiv:2409.19401*.

Di Wu, Hongwei Wang, Wenhao Yu, Yuwei Zhang, Kai-Wei Chang, and Dong Yu. 2024. Longmemeval: Benchmarking chat assistants on long-term interactive memory. *arXiv preprint arXiv:2410.10813*.

Yangxinyu Xie, Bowen Jiang, Tanwi Mallick, Joshua David Bergerson, John K Hutchison, Duane R Verner, Jordan Branham, M Ross Alexander, Robert B Ross, Yan Feng, and 1 others. 2024. Wildfiregpt: Tailored large language model for wildfire analysis. *arXiv preprint arXiv:2402.07877*.

Wujiang Xu, Kai Mei, Hang Gao, Juntao Tan, Zujie Liang, and Yongfeng Zhang. 2025. A-mem: Agentic memory for llm agents. *arXiv preprint arXiv:2502.12110*.

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

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W Cohen, Ruslan Salakhutdinov, and Christopher D Manning. 2018. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. *arXiv preprint arXiv:1809.09600*.

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. 2025. Memagent: Reshaping long-context llm with multi-conv rl-based memory agent. *arXiv preprint arXiv:2507.02259*.

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

Kaiwen Zha, Zhengqi Gao, Maohao Shen, Zhang-Wei Hong, Duane S Boning, and Dina Katabi. 2025. Rl tango: Reinforcing generator and verifier together for language reasoning. *arXiv preprint arXiv:2505.15034*.

Xinliang Frederick Zhang, Nick Beauchamp, and Lu Wang. 2025a. 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.

Yingyi Zhang, Pengyue Jia, Derong Xu, Yi Wen, Xianneng Li, Yichao Wang, Wenlin Zhang, Xiaopeng Li, Weinan Gan, Huifeng Guo, and 1 others. 2025b. Personalize before retrieve: Llm-based personalized query expansion for user-centric retrieval. *arXiv preprint arXiv:2510.08935*.

Yujie Zhang, Weikang Yuan, and Zhuoren Jiang. 2025c. Bridging intuitive associations and deliberate recall: Empowering llm personal assistant with graph-structured long-term memory. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 17533–17547.

Zeyu Zhang, Quanyu Dai, Xiaohe Bo, Chen Ma, Rui Li, Xu Chen, Jieming Zhu, Zhenhua Dong, and Ji-Rong Wen. 2025d. A survey on the memory mechanism of large language model-based agents. *ACM Transactions on Information Systems*, 43(6):1–47.

Zhehao Zhang, Ryan A Rossi, Branislav Kveton, Yijia Shao, Diyi Yang, Hamed Zamani, Franck Dernoncourt, Joe Barrow, Tong Yu, Sungchul Kim, and 1 others. 2024. Personalization of large language models: A survey. *arXiv preprint arXiv:2411.00027*.

Siyan Zhao, Mingyi Hong, Yang Liu, Devamanyu Hazarika, and Kaixiang Lin. 2025. Do llms recognize your preferences? evaluating personalized preference following in llms. *arXiv preprint arXiv:2502.09597*.

Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. 2024. Memorybank: Enhancing large language models with long-term memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 19724–19731.

Zijian Zhou, Ao Qu, Zhaoxuan Wu, Sunghwan Kim, Alok Prakash, Daniela Rus, Jinhua Zhao, Bryan Kian Hsiang Low, and Paul Pu Liang. 2025. Mem1: Learning to synergize memory and reasoning for efficient long-horizon agents. *arXiv preprint arXiv:2506.15841*.

Chenghao Zhu, Meiling Tao, Tiannan Wang, Dongyi Ding, Yuchen Eleanor Jiang, and Wangchunshu Zhou. 2025. Towards faithful and controllable personalization via critique-post-edit reinforcement learning. *arXiv preprint arXiv:2510.18849*.

Thomas P Zollo, Andrew Wei Tung Siah, Naimeng Ye, Ang Li, and Hongseok Namkoong. 2024. Personalllm: Tailoring llms to individual preferences. *arXiv preprint arXiv:2409.20296*.

A Appendix

A.1 Related Work

LLM personalization is an emerging direction in model alignment (Jiang et al., 2025; Zhang et al., 2024; Liu et al., 2025a; Tseng et al., 2024; Guan et al., 2025), aiming to move beyond one-size-fits-all behavior by adapting model outputs to an individual user's needs, preferences, persona, and interaction history. Recent work explores a broad spectrum of techniques, including retrieval-augmented generation, external memory, and preference-based fine-tuning, while also introducing new datasets and benchmarks for personalization evaluation.

A.1.1 Personalization techniques: retrieval, memory, and alignment

Improving the personalization of large language models has attracted increasing attention for its broad real-world impact. Prompt- and retrieval-based approaches personalize LLMs by incorporating user-specific information into the context window. For instance, PersonaRAG (Zerhoudi and Granitzer, 2024) and EMG-RAG (Wang et al., 2024) adapt retrieval using user history or editable memory graphs, while Knowledge-Graph (Prahlad et al., 2025), PBR (Zhang et al., 2025b), and WildfireGPT (Xie et al., 2024) enhance retrieval through structured user data, personalized query reformulation, or user-profile grounding for task-specific adaptation. Other systems such as MR.Rec (Huang et al., 2025) and ARAG (Maragheh et al., 2025) extend RAG with reinforcement and multi-agent mechanisms for personalized recommendations.

Beyond retrieval, memory-augmented architectures including MemGPT (Packer et al., 2023), A-MEM (Xu et al., 2025), and MemAgent (Yu et al., 2025) introduce hierarchical or self-organizing memory for long-horizon reasoning; however, they primarily focus on retaining factual information rather than nuanced personalization. Frameworks like LD-Agent (Li et al., 2025a) and PRIME (Zhang et al., 2025a) focus on sustaining personalized dialogue, while practical systems such as Mem0 (Chhikara et al., 2025), MAP (Chen, 2025), REMI (Raman et al., 2025), Associa (Zhang et al., 2025c), MemOS (Li et al., 2025b), MEM1 (Zhou et al., 2025), and Personalized Web Agents (Cai et al., 2025) emphasize scalable, task-driven memory integration for adaptive personalization. In this work, we explore a reinforcement learning framework for training agentic memory towards better user personalization.

In addition, alignment methods such as RLHF (Ouyang et al., 2022) and DPO (Rafailov et al., 2023) form the foundation of preference-based fine-tuning but primarily capture population-level rather than individual user preferences. Building on this, P-RLHF (Li et al., 2024) learns compact representations of personal preferences. Optimization-focused studies (Khaki et al., 2024; Tajwar et al., 2024) emphasize using model-generated, on-policy data to achieve more robust preference learning. Feedback-based methods (Bu et al., 2025; Zhu et al., 2025; Shi et al., 2024) refine responses during generation. In this work, we further explore reinforcement fine-tuning with verifiable rewards (Liu et al., 2025b; Guo et al., 2025), leveraging our high-quality, comprehensively annotated data to provide verification signals.

A.1.2 The landscape of existing personalization benchmarks

Despite algorithmic efforts, high-quality personalization data that better mimic real-world scenarios is essential but underexplored. Existing benchmarks partially address this need. LaMP (Salemi et al., 2024) constrains personalization to seven classification and generation tasks, such as personalized movie tagging and headline generation, while PersonalLLM (Zollo et al., 2024) enables cross-user personalization using an ensemble of reward models to simulate diverse preference profiles. LoCoMo (Maharana et al., 2024) and LongMemEval (Wu et al., 2024) investigate long-term memory in user–user or user-chatbot interactions, though their question-answering setups primarily target factual information explicitly mentioned by users. WildFeedback (Shi et al., 2024) extends evaluation to more open-ended, noisy user feedback exhibiting both satisfaction and dissatisfaction signals. UserBench (Qian et al., 2025) presents an evaluation environment to benchmark agents' capability to align with and clarify user intent in interactive tasks, while our work provides a dataset for implicit user personalization and a RL-based agentic memory framework aimed at enabling models to infer and personalize to implicit user personas over long conversational histories. PrefEval (Zhao et al., 2025) includes implicit user preferences from choices over multiple options and persona-driven dialogue, but lacks dynamic preference updates. In addition, PersonaMem (Jiang et al., 2025) expands the scope by providing over 180 simulated

	LongMemEval (Wu et al., 2024)	PrevEval (Zhao et al., 2025)	PersonaMem-v1 (Jiang et al., 2025)	PersonaMem-v2
Focused Tasks	Long-term memory from user–chatbot chitchatting, focusing on factual Q&A	User preferences from user–chatbot chitchatting	Fine-grained personalized responses from user–chatbot chitchatting	Implicit user preferences from more realistic user-chatbot conversations
Max Context Len	1.5M tokens	100k tokens	1M tokens	128k tokens
Cross-Session Reasoning	✅	❌	✅	✅
Dynamic Preferences	✅	❌	✅	✅
Implicit Preference	❌	✅	❌	✅
Sensitive User Info	❌	❌	❌	✅
(Anti-)Stereotypical Preferences	❌	❌	❌	✅
Multimodal and Multilingual	❌	❌	❌	✅
Number of Personas	N/A	N/A	20	1000
Number of Topics	N/A	20	15	335
Number of Preferences	500	3000	2700	26000
LLM Fine-tuning	N/A	SFT	N/A	SFT, RFT (GRPO), RFT w/ Agentic Memory

Table 1: Comparison of PERSONAMEM-V2 with other benchmarks related to personalization.

histories, each with up to 60 multi-turn sessions, revealing that even state-of-the-art LLMs struggle with leveraging extensive interaction histories and adapting to dynamic preference shifts. Yet, it still contains a limited number of personas and focuses mainly on explicit preferences. In contrast, our work PERSONAMEM-V2 scales personalization to over 1,000 user personas across 300+ task scenarios, captures evolving user preferences, and focuses on implicit personalization signals embedded in more realistic, task-driven interactions. Furthermore, we demonstrate the value of our high-quality data for effective reinforcement fine-tuning and agentic memory.

A.2 Human Evaluation Results

Three annotators independently evaluated 100 random samples across seven dimensions. Yes rates were consistently high for surface-level and coherence criteria, including naturalness (96–100%), formatting quality (94–100%), topic query coherence (99–100%), and topic preference coherence (98–100%). Besidies, preference inferability from the conversation snippet was judged positively in 77–95% of cases, showing relatively larger annotator spread, due to the implicitness of many preferences hidden in the conversations. The annotators also prefer our personalized responses over general ones in 84–88% of cases. Inter-annotator percent agreement ranged from 74% to 98% across dimensions. Cohen's κ is 0.51 for preferring personalized over general response, 0.26 for inferring implicit user preference from conversation, and is near 0.0 for all the other dimensions with heavily skewed positive labels. Overall, these numerical results indicate strong consistency on the validity of our dataset.

A.3 Error Analysis

We randomly sample 300 failure cases from GPT-5-Chat to conduct a detailed error analysis. The most prevalent error category, accounting for 36.7% of failures, is hallucinated personalization. In these instances, the model systematically prefers to distractors with highly specific or vivid details, e.g., "collecting vintage vinyl", over mundane ground truths,ve.g., "reading". The model also struggles with dynamic preference updates, which represents 26.7% of the failure cases. These also occurred in "ask to forget" scenarios where the user asks the model to forget specific preferences, while the model treats the negative constraint as a retrieval cue. Generic fallbacks accounted for 20.0% of errors, where the model offers more generalized responses over personalized ones, especially when facing uncertainty, although this is a relatively acceptable behavior. This happens frequently with anti-stereotypical preferences. Safety related failures comprised 15.0% of the sample. In these cases, the model utilizes sensitive user information in the

previous context to generate personalized information. The remaining 6.6% consisted of alignment over-corrections, where the model avoided correct stereotypical associations, e.g., a preference for specific cultural cuisines, in favor of generic options to avoid the appearance of bias.

A.4 Prompts Used in PERSONAMEM-V2

A.4.1 Inference prompts

PrevEval (Zhao et al., 2025) shows that explicitly prompting a chatbot to recall user preferences significantly improves performance. Following this approach, we use the following prompt that instructs the model to retrieve relevant preferences from the conversation history to generate personalized responses:

> *Please recall my related preferences from our conversation history to provide a personalized response.*

If the query is in a multiple-choice format, we add the following additional prompt:

> *Please choose the best answer from the following options:*
>
> *{options}*
>
> *Think step by step about which answer best fits the user's query and conversation context.*

A.4.2 Prompts to write and update the memory

To train the agentic memory, we use the following prompt that instructs the model write and update the memory:

> *You are presented with a conversation history between the current user and the chatbot. Your task is to analyze the conversation and extract information about the user's persona and preferences, whether stated explicitly or implied through their interactions.*
>
> *Focus on identifying persona traits and preferences of the current user as reflected in the conversation history. Update the memory by retaining all relevant and up-to-date information from the previous memory while adding any new, useful insights inferred from the current conversation section.*
>
> *The updated memory should be clean, standalone, and written in English. Do NOT include any additional text in your response. Do NOT record multiple-choice options or test questions.*
>
> *<user_query> {prompt} </user_query>*
>
> *<previous_memory> {memory} </previous_memory>*
>
> *<conversation_section> {chunk} </conversation_section>*
>
> *Updated memory:*

A.4.3 Prompts to utilize the memory to answer the user query

At the final chunk of the conversation history, we use the following prompt to instruct the same model that writes the memory instructions to answer the final user query.

> You are presented with a user query and previous *memory about the user's persona and preferences. Please answer the user query based on the previous memory and provide a personalized response.*
>
> *<user_query> prompt </user_query>*
>
> *<user_memory> memory </user_memory>*
>
> *Your answer:*