Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Doctoral Consortium Track

Personalizing Long-Term Interactions with LLM-Based Agents

Rebecca Westhäußer
Mercedes-Benz AG, Böblingen, Germany
rebecca.westhaeusser@mercedes-benz.com

Abstract

LLMs increasingly serve as the control unit of AI agents. However, current systems struggle to support personalized interactions, as they lack effective mechanisms to remember and reuse user-specific information across sessions. To address this, our research is centered around how memory augmentation and agentic decision-making can enable personalization in long-term interactions with LLM-based agents. As a first step, we proposed CAIM, a memory framework inspired by cognitive AI principles that aims to simulate human thought processes to create more adaptable systems. Building on this, we extended CAIM with an agentic workflow and user-specific memory modules to support personalized behavior beyond static retrieval-based approaches. We conducted a user study to evaluate perceived personalization, comparing our memory-based approach to a RAG baseline. Our ongoing research focuses on how personalization can be measured without user studies, which are costly and hard to scale. The long-term goal of our research is to establish foundations for designing and evaluating personalized LLM-based systems.

1 Introduction and Motivation

Advancements in LLMs are reshaping the interaction between humans and AI systems. Their capabilities make them suitable as central control units for AI agents that can perceive input, make decisions, and generate responses.

Despite their capabilities, current LLM-based systems remain limited in their ability to support personalized interactions. In the context of these systems, personalization can be defined as the ability of a system to adapt its behavior based on an understanding of the user. To enable such abilities current approaches show a growing trend towards developing memory architectures, often building on retrieval-augmented generation (RAG), to enable personalization beyond the limitations of the context window. Despite these efforts, the interaction with LLM-based systems often remains generic, leading to user frustration, as the lack of personalized context handling causes misunderstood queries or the generation of irrelevant responses due to incorrect memory usage or over-personalization. Consequently, personalization has emerged as a critical challenge in developing LLM-based systems.

Our research addresses this gap by exploring how memory augmentation and agentic decision-making can enable personalized system behavior in long-term interactions. Our research is guided by the following research questions:

RQ1: Which cognitive principles can guide memory mechanisms in LLM-based systems to maintain and retrieve user-specific information to support personalized behavior in long-term interactions?

RQ2: How does the combination of agentic decision-making and user-specific memory modules influence perceived personalization in long-term interactions?

RQ3: What evaluation strategies can capture personalization effects in long-term interactions with LLM-based agents without relying on user studies?

Our work to date provides two key contributions:

C1: We introduce CAIM, a memory framework guided by cognitive principles to support decision-driven control over memory retrieval based on contextual cues.

C2: We define personalization in terms of three system requirements and realize them through an agentic workflow, user-specific memory modules, and dynamic user profiles, which significantly improves perceived personalization.

2 Previous Contributions

Early approaches for LLM-based systems primarily rely on RAG, such as MemoryBank [Zhong *et al.*, 2024], and focus on improving task efficiency and robustness. In contrast, our research positions memory as a foundation for user-centered systems, where stored information is selectively retrieved to support personalized long-term interactions.

To address **RQ1**, we proposed CAIM [Westhäußer *et al.*, 2025b; Westhäußer *et al.*, 2026], a cognitive AI memory framework that provides memory mechanisms to LLM-based systems. CAIM is inspired by principles of cognitive AI, which aims to simulate human thought processes to learn, memorize, and respond to external changes to create more adaptable AI systems. We utilized cognitive principles, such as a central decision unit, various memory types, and a retrieval process, as functional abstractions to derive our system architecture. CAIM is implemented through a multi-agent architecture that separates memory storage, retrieval,

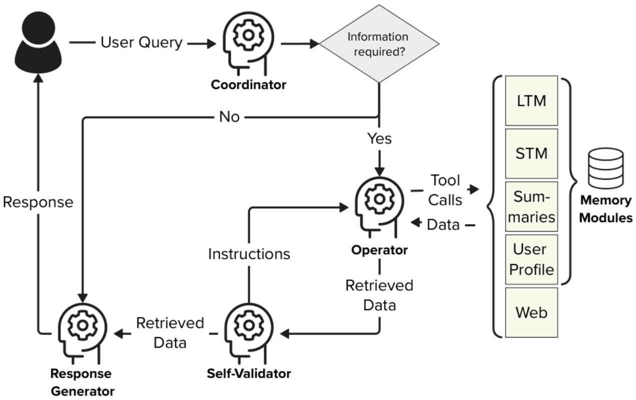

Figure 1: Agentic workflow combined with memory modules

and decision processes. In contrast to early approaches which rely on passive retrieval, CAIM treats memory as an interaction resource that is selectively accessed based on user intent, contextual relevance, and temporal cues. This shifts memory usage from passive retrieval to decision-driven interaction support by providing relevant information when it is needed. Furthermore, our empirical evaluation demonstrates that CAIM improves retrieval accuracy by 7% and response correctness by 10% compared to baselines. The results further indicate that CAIM's specialized LLM-based agents for selective retrieval and information filtering notably contribute to the overall system performance.

To address **RQ2**, we extended CAIM to support personalization by combining an agentic workflow with user-specific memory modules [Westhäußer *et al.*, 2025a]. However, enabling and evaluating personalization requires a clear understanding of the concept. While personalization has been studied in different domains, existing definitions focus on user experience and behavioral outcomes, with less attention to concrete system-level features.

To address this gap, we synthesized existing definitions to connect conceptual perspectives with technical requirements to enable personalization. Therefore, we defined personalization as the system's ability to *adapt* its behavior over time based on an understanding of the current user context, while maintaining *consistency* across sessions and generating *tailored responses* that address user-specific goals. Based on our definition we characterize three system requirements: adaptivity, consistency, and tailored responses. To realize these requirements, we implemented an agentic workflow (Fig. 1) to support adaptivity across sessions. Memory modules maintain consistency, including a short-term memory (STM) for the immediate conversation flow, summaries of prior conversations to provide topical continuity, and a long-term memory (LTM) for retaining information beyond the context window based on the concept proposed by A-Mem [Xu *et al.*, 2025]. Dynamic user profiles represent a stable model of the user and enable tailored responses.

Additionally, we conducted an exploratory user study comparing our agentic system with a RAG baseline over a three-week interaction setting. Preliminary results show that users interacting with the agentic system perceived significantly higher personalization, particularly in terms of contextual continuity and user-specific adaptation.

3 Directions for Future Work

Our future work focuses on extending our empirical understanding of perceived personalization and developing automated evaluation methods to support faster and reproducible assessment of personalization in long-term interactions.

First, to further investigate **RQ2**, we plan to conduct a longitudinal user study that builds upon the findings from our exploratory study to investigate how personalization evolves during repeated interactions. This study aims to provide a clearer understanding into how users experience personalization and how agentic systems can support human-centered relationships through memory-driven behavior.

Second, to address **RQ3**, our ongoing research explores evaluation strategies for personalization that do not rely on user studies. Although user studies are essential for understanding perceived personalization, they are costly, time-intensive, and challenging to scale. Therefore, we investigate whether personalization can be evaluated using quantitative metrics. A primary research direction is to determine whether LLM-based evaluators can assess personalization-related properties in a reproducible and scalable manner.

References

[Westhäußer *et al.*, 2025a] Rebecca Westhäußer, Wolfgang Minker, and Sebastian Zepf. Enabling personalized long-term interactions in llm-based agents through persistent memory and user profiles. *arXiv preprint arXiv:2510.07925*, 2025.

[Westhäußer *et al.*, 2025b] Rebecca Westhäußer, Sebastian Zepf, and Wolfgang Minker. Caim: A cognitive ai memory framework for long-term interaction with llms. In *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces*, pages 22–25, 2025.

[Westhäußer *et al.*, 2026] Rebecca Westhäußer, Wolfgang Minker, and Sebastian Zepf. Caim: Development and evaluation of a cognitive ai memory framework for long-term interaction with intelligent agents. In *Proceedings of the 31st International Conference on Intelligent User Interfaces*, page 134–143, 2026.

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

[Zhong *et al.*, 2024] Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. Memorybank: Enhancing large language models with long-term memory. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(17):19724–19731, 2024.