Users' Expectations and Practices with Agent Memory

Brennan Jones[1], Kelsey Stemmler[1], Emily Su[1], Young-Ho Kim[2], Anastasia Kuzminykh[1]

brennan.jones@utoronto.ca, kelsey@stemmler.tech, em.su@mail.utoronto.ca, yghokim@younghokim.net, anastasia.kuzminykh@utoronto.ca

[1]Faculty of Information (iSchool), University of Toronto [2]NAVER AI Lab

Thanks for your interest!

Paper *Video*

Motivation

- AI agents and chatbots are becoming more prevalent in people's daily lives, serving tasks like writing feedback, brainstorming, general advice, and code debugging.
- Such AI tools have the potential to work in collaborative complementarity with the user – through the user and the system achieving co-understanding with each other.
- However, to achieve sufficient co-understanding in longer-term interactions, the agent needs to remember the right information about the user, and the user needs to understand the agent's memories and how they impact its actions.
- There is a lack of understanding of **users' expectations of AI systems' memory mechanisms** over longer periods of time, and of **users' practices of managing these systems' memories**.

Research Questions

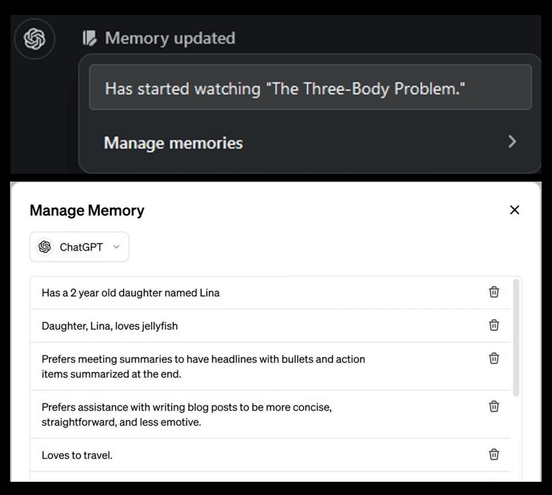
1. What are people's **mental models and expectations** of AI memory?
2. What are people's **unsatisfied needs** with AI memory?
3. What are people's **current practices** with managing and exploring the memory of a personalized AI system?

Data Collection and Analysis

- We conducted **one-on-one interviews with six participants** who use personalized AI tools with long-term memory (e.g., ChatGPT, Claude, NotebookLM, Replika) on a regular basis.
- We also **collected a series of discussion threads on Reddit**, in which people discussed their experiences using the memory features of personalized AI systems.
- We **thematically coded** the interviews together with the Reddit discussions, to understand people's expectations, practices, and pain points with the memory features of these systems.

FINDINGS

Finding 1: Users have an **incomplete understanding of AI agents' memory mechanisms**, including:



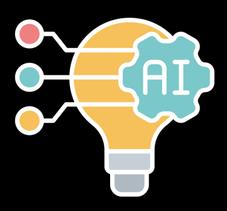
(1) Incomplete awareness of **what agents remember** in the short and long term:

*"When I'm conversing in it [...] it'll also sometimes say 'memory updated'. But, I think **I don't really know how it decides what is important information to remember versus what is not important information to remember** in, like the prompts that I write."* - P3

*"I didn't notice it, but now I'm looking at the memory, it's definitely collected a bunch of information about me. [...] **It's scary.**"* - P1

*"So it saved my supervisor['s name], which makes sense, because it reads emails, like, I polish my emails to my supervisor with GPT, so it knows. **And the really scary one is, it has my student ID, and my employee ID.** I think I don't put [those] there a lot. [...] but [ChatGPT's memory] captured it."* - P6

(2) Incomplete awareness of how memories **influence the agent's <u>behaviours and outputs</u>:**

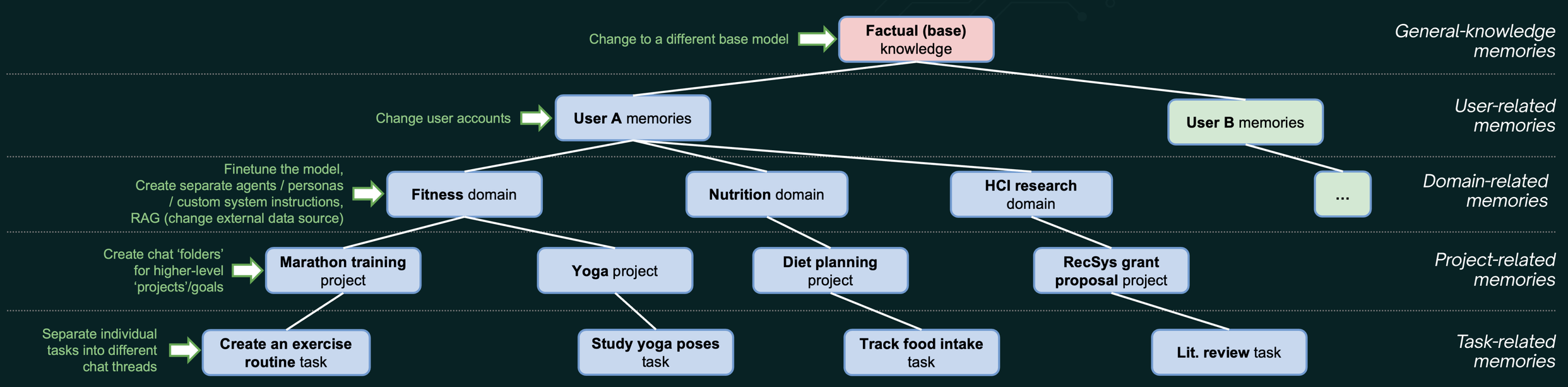


*"It would be interesting, like I mentioned, **to know which parts of the memory it has about me go into, like, which responses**, and to understand more about how it decides which portions [...] of that list [of memories] is like... you know, [it is] 100 or 200 items... like, that's a lot of things to remember about me. And to decide, like, which parts of that list actually go into a response versus not."* - P3

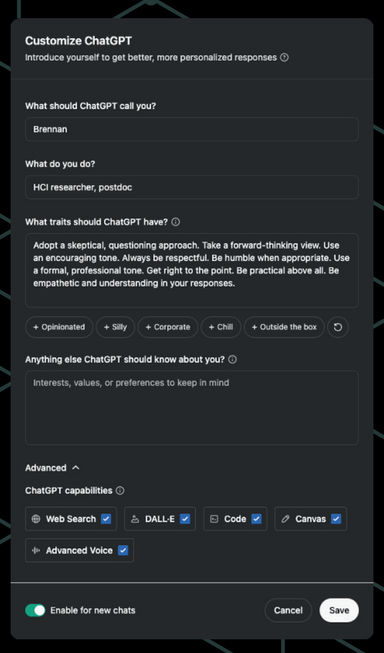
Finding 2: Heterogeneity of **Memory Needs** and **Management Practices:**

Users **do not consider all types of agent memory to be equal**, and **users' needs for memory recall evolve over time**, from task to task.

*"Let's say, **I don't want my conversation about research to overlap with a side project conversation**, but I might still be using the same information. And so, just for convenience or visual representation, I open two separate chat windows for it. And in that case I would prefer it if the agent preserved those memories. **But I also think of it as something that the user should have more control over.**"* - P4



Finding 3: Pain Points with Existing Memory Management Approaches:

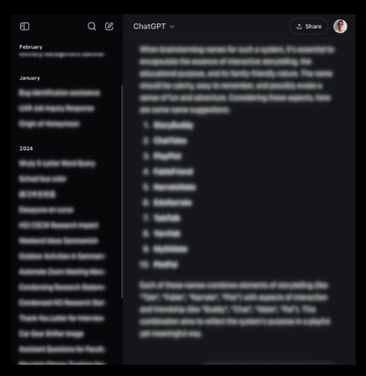


Users need to **manually 'pass on' memories from one session or agent to another** when switching tasks, projects, or domains.

- E.g., needing to **copy and paste the base/system instructions** (for system role, etc.) every time a new chat session is created.
- E.g., needing to **copy and paste relevant context information** from one chat session to another.

Users want the system to **remember some, but not all information** across conversations and sessions.

- E.g., users want personal and sensitive information to be stored and recalled only to the degree necessary for the task.

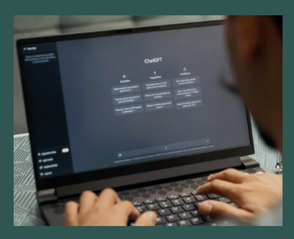


*"If it's not able to recall certain discussions that I had with [it], then I would try to like, feed again, [...] like, dump data, whatever I have. **Or maybe copy and paste the data from the previous chat [...] copy, paste the same conversation to it again.**"* - P5

*"I would say, it's kind of difficult to do now. **The only way I can keep it consistent is to copy and paste the same prompt.**"* - P2

*"It would be helpful if I could have like, you know... like, **different settings, or like different fields that I can customize**, right? Like, say, I want, like, a UX expert. Then I can have that as the standard, and then just swap out, like, other things."* - P3

Design Opportunities



Design memory mechanisms to **recognize different stages of a user's tasks**, and **abstract access to information based on the user's current task**.

- Implicitly recognize task or activity **as a context cue for memory retrieval**.
- Provide the user with **mechanisms for organizing memories into groups**, and setting the system's level of access to memories within these groups.

CHI 2025
April 26–May 1, 2025 in Yokohama, Japan