File 2610.34073, filed 25 October 2026, section: Context & memory

Memory Tax: Dissecting the Net Value of Agent Memory

Xuan He, Sining Zhoubian, Ruozhen Yang, Yuhao Wen, Zhineng Wang, Yuyang Hu, Ke Yang, Yuyu Zhang et al.. UIUC

In brief

A large evaluation finds that many dedicated agent memory systems underperform naive RAG over raw history. The authors call the losses a 'memory tax'.

Why builders should care

Before you build a memory layer, you should know if it beats just retrieving the raw history. Often it does not.

The paper, in plain English

Memory is increasingly used in long-horizon agents, but it has been unclear whether building memory beats simply retrieving from the raw history.

Across multiple benchmarks and model backbones, the authors find that many dedicated memory systems underperform naive RAG in many cases, a negative net value.

Using RAG as the reference, they split the net value of memory into gains (memory benefit) and losses (memory tax). Losses arise because memory processing reorganizes or distills information and can remove or obscure evidence the task needs. They study how memory tax arises from both system and evaluation choices, and plan to open-source code and results.

How to apply it

1. Make plain RAG over raw history your baseline. If your memory system does not beat it, ship RAG.
2. For every memory transformation (summarizing, distilling preferences), check what evidence it discards.
3. Keep raw history available as a fallback the agent can consult.
4. Evaluate memory on net value, not standalone scores.

Key results

Quoted from the abstract

> *Through a large-scale evaluation across multiple benchmarks and model backbones, we find many dedicated memory systems underperform naive RAG in many cases, indicating a negative net value.*

> *Taking RAG as a reference, we dissect the net value of memory into the gains and losses over direct retrieval, termed memory benefit and memory tax.*

Summary written from the arXiv abstract. Read the original: https://arxiv.org/abs/2610.34073

Curated by Atharv Pawar