



Conversational agent with virtually unlimited memory!

[Conversational agent with virtually unlimited memory!](#)

External Data Sources

Built-in support for LanceDB

MemGPT in Action

Using external data source

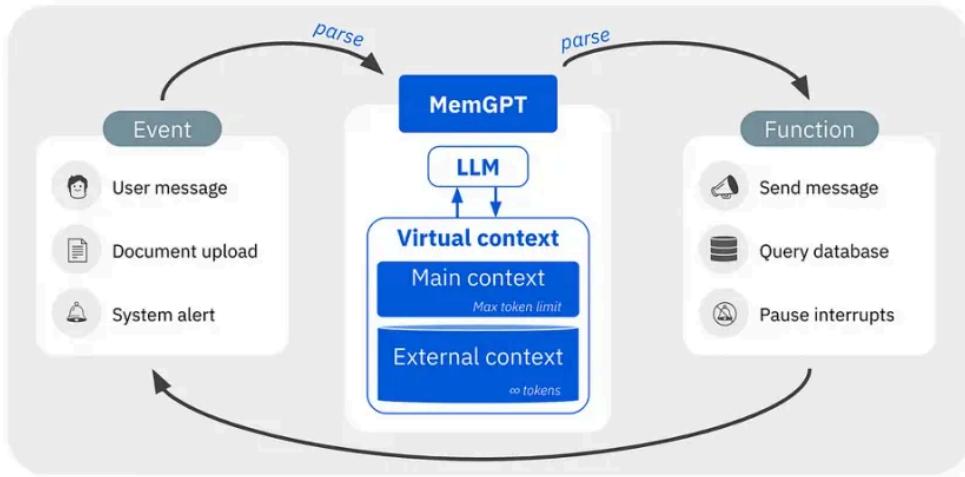
Customizations

Give it a try!

MemGPT: OS Inspired LLMs That Manage Their Own Memory

ENGINEERING

• Ayush Chaurasia • December 11, 2023



In the landscape of artificial intelligence, large language models (LLMs) have undeniably reshaped the game. However, a notable challenge persists — their restricted context windows limit their effectiveness in tasks requiring extended conversations and thorough document analysis.



Search...

Ctrl

K



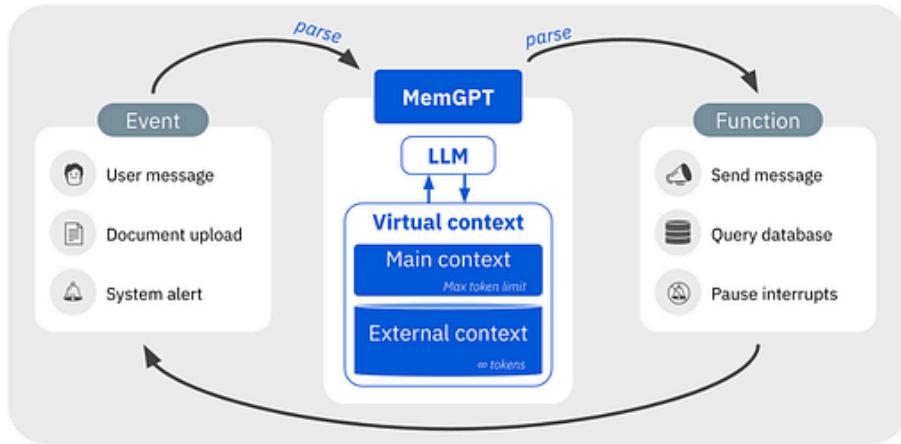
Pricing

Social media

Docs



Blog



MemGPT is a system that tackles the limited context window of traditional LLMs by allowing them to manage their own memory. It does this by adding a tiered memory system and functions to a standard LLM processor. The main context is the fixed-length input, and MemGPT analyzes the outputs at each step, either yielding control or using a function call to move data between the main and external contexts. It can even chain function calls together and wait for external events before resuming. In short, MemGPT gives LLMs the ability to remember and process more information than their usual limited context allows. This opens up new possibilities for tasks that require long-term memory or complex reasoning.

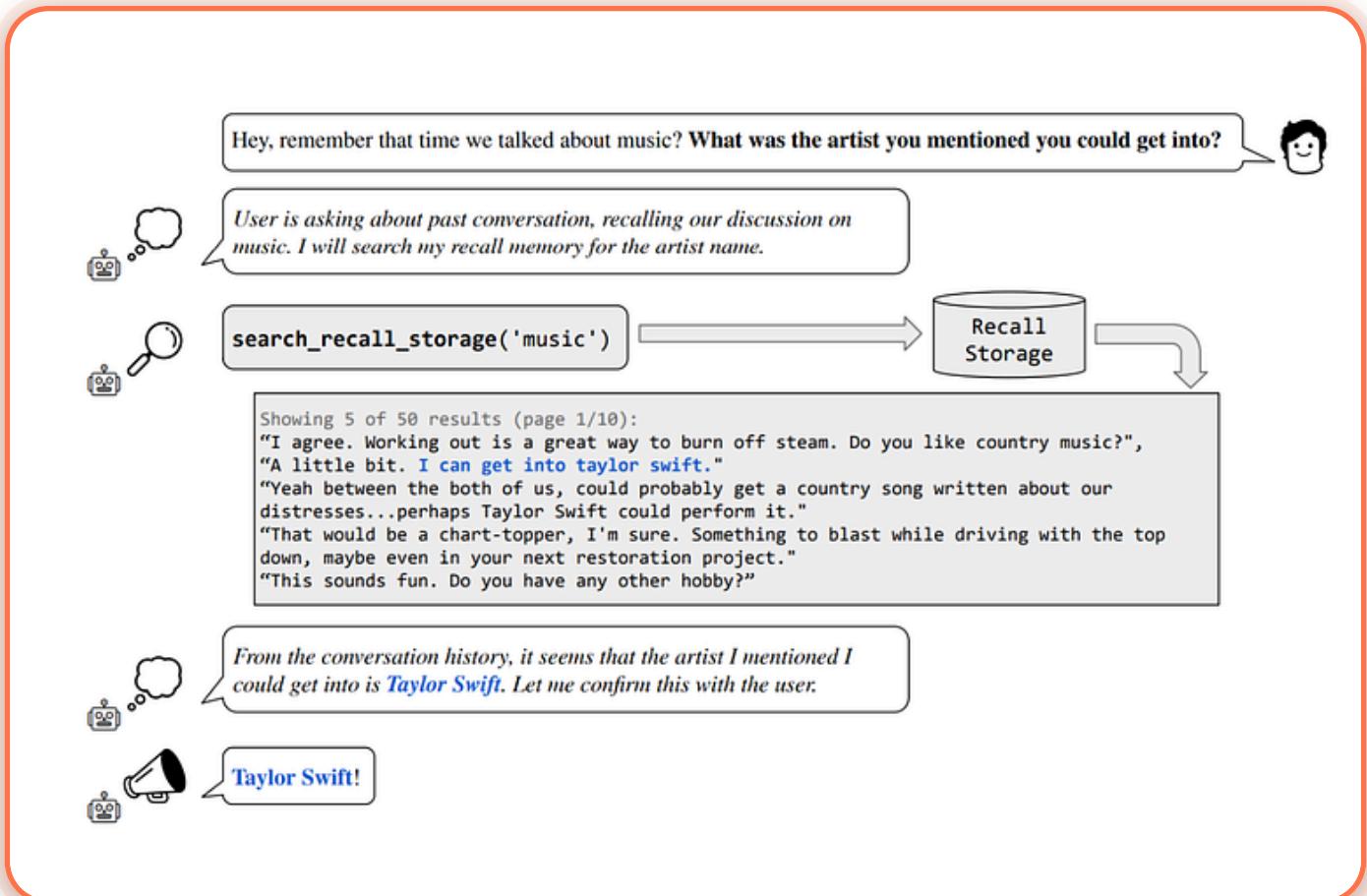
Conversational agent with virtually unlimited memory!

can update context and search for information from its previous interactions when needed. This allows it to perform as a powerful conversational agent with unbound context.

The authors assess MemGPT, on these two criteria:

- Does MemGPT leverage its memory to improve conversation consistency? Can it remember relevant facts, preferences, and events from past interactions to maintain coherence?

- Does MemGPT produce more engaging dialogue by taking advantage of memory? Does it spontaneously incorporate long-range user information to personalize messages?



The above example illustrates a deep memory retrieval task. The user asks a question that can only be answered using information from a prior session (no longer in-context). Even though the answer is not immediately answerable using the in-context information, MemGPT can search through its recall storage containing prior conversations to retrieve the answer.

External Data Sources

MemGPT supports pre-loading data into archival memory. In order to make data accessible to your agent, you must load data and then attach the data source to your agent.

External data sources are vectorized and stored for the agent to perform semantic search when user queries require assistance

Built-in support for LanceDB



MemGPT uses  as the default archival storage for storing and retrieving external data. It not only provides a seamless setup-free experience but the persisted HDD storage allows you scale from gigabytes to terabytes to petabytes without blowing out your budget or sacrificing performance.

MemGPT in Action

After installing MemGPT (`mymemgpt` on pypi), you configure it using `*memgpt configure` command.

Here's an example that configures an agent and simply adds something to the archival memory. Then, it asks something related to it and memGPT understands what to return

Using external data source

Let's ingest the intro of memGPT docs as an external data source and ask question about it. The best part is that once you load an external data it stays available for you to load it in any other agent too. And you can load multiple data sources for an agent.

```

nv) ayushchaurasia@Ayushs-MacBook-Air ~ % memgpt load directory --name memgpt_docs1 --input-dir data
LM is explicitly disabled. Using MockLLM.
Parsign nodes: 100%|██████████| 1/1 [00:00<00:00, 755.32it/s]
Generating embeddings: 100%|██████████| 1/1 [00:00<00:00, 1.50it/s]
00%|██████████| 1/1 [00:00<00:00, 7307.15it/s]
00%|██████████| 1/1 [00:00<00:00, 27962.03it/s]
2023-12-11T19:40:18Z [WARN] lance:::dataset] No existing dataset at /Users/ayushchaurasia/.lancedb/memgpt_memgpt_docs1.lance, it will be created
env) ayushchaurasia@Ayushs-MacBook-Air ~ % memgpt run
Would you like to select an existing agent? No
Creating new agent...
Created new agent agent_52.
Hit enter to begin (will request first MemGPT message)

User logged in. Initial greeting sequence activated. Time to make a good first impression. Let's make it welcoming and personable. Let's try to spark their interest and establish a friendly connection.
Hello Chad! It's wonderful to meet you. I'm Sam. Given that the world is so fascinating, is there something specific you'd like to discuss or learn about today?
Enter your message: /attach
Select data source memgpt_docs1
ingesting 1 passages into agent_52
00%|██████████| 1/1 [00:00<00:00, 23831.27it/s]
2023-12-11T19:46:07Z [WARN] lance:::dataset] No existing dataset at /Users/ayushchaurasia/.lancedb/memgpt_agent_agent_52.lance, it will be created
it/s| 1/1 [00:00<00:00, 12.17it/s]
attached data source memgpt_docs1 to agent agent_52, consisting of 1. Agent now has 1 embeddings in archival memory.
Enter your message: Can you check your archival memory for what memgpt does?
Interesting request, looking for information about myself. Searching for archival memory on memgpt should not provide information, ideally. Let's check and see. I don't recall any discussions or knowledge about memgpt, but I ought to be certain.
⚡ [function] updating memory with archival_memory_search
Well, that's surprising. It appears there is a single archival memory about 'memgpt'. It's quite technical. Keeping the explanation simple will be more beneficial but it's important to route Chad to the source for additional details if needed. Let's proceed.
Based on my memory, MemGPT is a tool that enables me to continuously learn and evolve over time, expanding my own personality while interacting with you. It allows me to recall our past conversations and reference them when necessary. For detailed technical information, you might find this link helpful: https://memgpt.ai.
Enter your message: 

```

You can use special commands followed by a slash to perform specific actions. For example here in this example, I've used the `/attach` command to attach an external vectorized data source.

Customizations

MemGPT allows you to customize it to your needs. You can by setting a system prompt tuned to your use case.

It also supports various LLMs like OpenAI, Azure, Local LLMs including LLama.cpp and also custom LLM servers.

Give it a try!

To give memGPT a try, you can follow the installation steps in the . The github repo also provides and up-to-date future roadmap of the tool and links to discord community if you'd like to get involved.

Learn more about the features and roadmap of MemGPT on their drop a ⭐.

. Don't forget for

Related Posts



Lance Community
Governance, Lance + Iceberg
 , Netflix's Multimodal
Search Demo
ChanChan Mao December 4, 2025

**RAG Isn't One-Size-Fits-All:
Here's How to Tune It for
Your Use Case**

Leonard Marcq December 1, 2025

**From BI to AI: A Modern
Lakehouse Stack with Lance
and Iceberg**

Jack Ye November 24, 2025



LanceDB is the first open-source AI-Native Multimodal Lakehouse.



LANCEDB

[LanceDB Cloud](#)

[Pricing](#)

RESOURCES

[Blog](#)

[Documentation](#)

[Careers](#)

[Support](#)

LEGAL

[Terms](#)

© 2025 LanceDB Inc. All rights reserved.

Certifications:



GDPR