



Coding Nexus · [Follow publication](#)

★ Member-only story

OpenMemory: The Open-Source Brain That Lets AI Remember

3 min read · 1 day ago



Algo Insights

Following

Listen

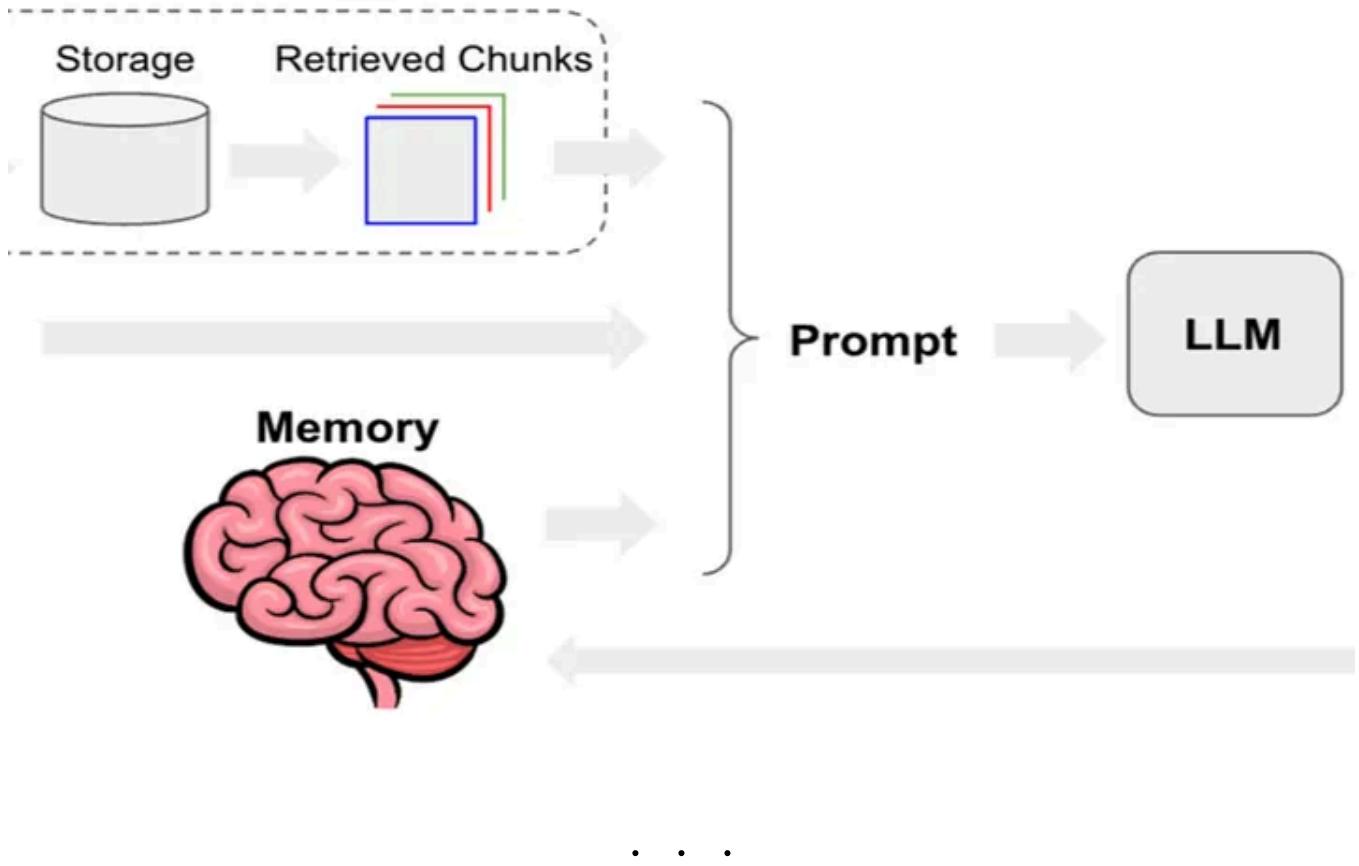
Share

More

Most AI “assistants” forget everything the moment you close the chat. They don’t remember your name, your preferences, or that one thing you told them yesterday.

That’s why I find **OpenMemory** quite interesting. It’s open-source, runs locally, and provides AI systems *with real memory* — similar to a brain for your chatbot or copilot.

Retrieval (Optional)



What Exactly Is OpenMemory?

Think of it like your AI's notepad — but smarter. It doesn't just store random sentences; it actually *understands* what you said and organises it by meaning.

For example, if you tell your bot:

“I like working in dark mode.”

That's not just a string of text — OpenMemory stores it as a *semantic memory* so the AI can recall it later when relevant.

It's built to remember things long-term — securely and privately — without depending on third-party services.

• • •

How It Works

Behind the scenes, OpenMemory uses a method called **Hierarchical Memory Decomposition**.

Don't worry, that's just a fancy way of saying it sorts information into categories, similar to how humans think.

Memory Type	What It Stores	Example
Episodic	Events	"User logged in at 9 AM."
Semantic	Facts	"User prefers dark mode."
Procedural	Actions	"Send reminder every morning."
Emotional	Feelings	"User seemed frustrated."
Reflective	Insights	"User is more productive in the morning."

So your AI isn't just searching through a huge list of notes — it's recalling information from the *appropriate type* of memory, making it quicker and more precise.

• • •

Why It's Different From the Rest

There are many “memory layers” available, but most are cloud-based and expensive. OpenMemory, by contrast, is open-source, runs locally, and provides you with full control over your data.

Here's the gist:

Feature	OpenMemory	Typical SaaS Memory
Open-source	Yes	No
Self-hosted	Yes	No
Fast recall (100k nodes)	~120 ms	~350 ms
Explainable memory links	Yes	No
Cost (per 1M tokens)	~\$0.40	~\$3.00+
Data ownership	100% yours	Vendor controlled

So not only is it faster and cheaper, but you're not sending private info to someone else's server.

Setting It Up

You can get OpenMemory running in a few minutes.

```
git clone https://github.com/caviraoss/openmemory.git
cd openmemory/backend
cp .env.example .env
npm install
npm run dev
```

It'll start a local API at

<http://localhost:8080>

Want to go production-grade? Just set it up with Docker:

```
docker compose up --build -d
```

That's all. No complicated cloud setup.

• • •

Let's Try It

Once it's running, you can add and recall memories using a few curl commands.

Add a memory:

```
curl -X POST http://localhost:8080/memory/add \
-H "Content-Type: application/json" \
-d '{"content": "User prefers dark mode"}'
```

Query what the AI remembers:

```
curl -X POST http://localhost:8080/memory/query \
-H "Content-Type: application/json" \
-d '{"query": "What color theme does the user like?"}'
```

And it'll respond with:

```
[  
  { "content": "User prefers dark mode", "score": 0.93 }  
]
```

It's that straightforward.

• • •

Works With LangGraph Too

If you're using LangGraph to create agentic workflows, OpenMemory integrates seamlessly.

You toggle several environment variables such as:

```
OM_MODE=langgraph  
OM_LG_NAMESPACE=default  
OM_LG_REFLECTIVE=true
```

Now each node (observe, plan, act, reflect) automatically links to its own “memory sector.”

That means your agent can *learn from its own thoughts* — bringing it closer to true intelligence.

• • •

What About Cost?

Here's what surprised me the most — you can run a reliable AI memory system for under \$10 a month on a VPS.

Metric	OpenMemory	SaaS Memory
Latency	~120 ms	~350 ms
Storage (1M memories)	~\$3/month	~\$60+
Monthly total	~\$5-8	~\$60-120

Pretty good deal, right?

• • •

Privacy Comes First

Everything remains local.

You can encrypt data, isolate users, and wipe it clean anytime you want. No vendor lock-in, no “trust us” data policies.

• • •

My Take

OpenMemory feels like a small but important step toward *genuine* agentic AI. It provides models with something they have always lacked—lasting **memory**.

If you're exploring copilots, journaling bots, or AI agents, give it a try.

You might finally have a system that recalls what you told it yesterday.

[GitHub: caviraoss/openmemory](#)

AI

Ai Agent

Llm

Llm Agent

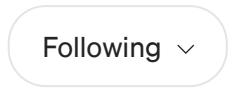
Ai Tools

Follow

Published in Coding Nexus

7.9K followers · Last published 1 hour ago

Coding Nexus is a community of developers, tech enthusiasts, and aspiring coders. Whether you're exploring the depths of Python, diving into data science, mastering web development, or staying updated on the latest trends in AI, Coding Nexus has something for you.

Following ▾

Written by Algo Insights

3.4K followers · 6 following

Algo Insights: Unlocking the Future of Trading with Code, Data, and Intelligence.

https://linktr.ee/Algo_Insights

No responses yet



Bgerby

What are your thoughts?

More from Algo Insights and Coding Nexus

 In Coding Nexus by Algo Insights

4 Open-Source Tools That Made Me Rethink My Dev Setup

I've been coding for a while now. Most of the tools we use every day... they've been the same for years. Editors, browsers, frameworks. Just...

 Sep 3  451  9



...

 In Coding Nexus by Civil Learning

The Guy Who Let ChatGPT Trade for Him—and Somehow It Worked

You know how everyone says, "Don't let AI touch your Money"? Well, someone on Reddit decided to ignore that.

 Oct 8  156  7



...

In Coding Nexus by Civil Learning

6 AI Tools You've Probably Never Tried (But Should)

Most people stop at ChatGPT. And honestly, that's fine—it's a fantastic tool.

 Sep 26  355  7



...

In Coding Nexus by Algo Insights

How to Build Your Own AI Rig for Running Local LLMs (Gemma, Mistral, Qwen, GPT-OSS and Llama)

About three months ago, I realised I was utterly dependent on companies that didn't care about anything except power, Money, and control.

star Oct 8 354 12



...

[See all from Algo Insights](#)

[See all from Coding Nexus](#)

Recommended from Medium

 In Towards Deep Learning by Sumit Pandey

Why Everyone Will Want DGX Spark on Their Desk—Yes, Everyone

I just saw this picture today and was amazed, I've been waiting for this moment for a long time. (No, it's not Elon.) It's that tiny...

 Oct 14  53  11



...

 In Coding Nexus by Civil Learning

MarkItDown: Convert Anything into Markdown—the Smart Way to Feed LLMs

You know that feeling when you're trying to feed a PDF or a Word document into an LLM, and it just doesn't understand what's going on...

◆ Oct 15 🙋 224 💬 4



In Towards AI by Gao Dalie (高達烈)

RAG is Not Dead! No Chunking, No Vectors, Just Vectorless to Get the Higher Accuracy

Over the past two years, I have written numerous articles on how Retrieval-Augmented Generation has become a standard feature in nearly all...

◆ Oct 17 🙋 534 💬 4





In Dare To Be Better by Max Petrusenko

Claude Skills: The \$3 Automation Secret That's Making Enterprise Teams Look Like Wizards

How a simple folder is replacing \$50K consultants and saving companies literal days of work



Oct 17



333



5



...



Bytefer

The Self-Hosted Alternative to Notebook LM/Perplexity Is Here.

NotebookLM and Perplexity Are Impressive, But They're Missing Half the Picture.



4d ago



98



2



...

 In AI Software Engineer by Joe Njenga

Why Claude Weekly Limits Are Making Everyone Angry (And \$100/Month Plan Will Not Save You)

Yesterday, I finally hit my weekly Claude limit, and I wasn't surprised, since I see dozens of other users online going crazy over these...

 Oct 19  123  23

[See more recommendations](#)