

# Mem0: Building Production-Ready AI Agents with Scalable Long-Term Memory

Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, and Deshraj Yadav

research@mem0.ai  
 开源代码: <https://github.com/mem0ai/mem0>  
 mem0

## 简介

- 本文为llm/llm-based agent设计了**两套**类似人类记忆机制的**memory**方案: 1) mem0, 其在每轮交互(query, response)后, 用llm进行本轮的**记忆信息提取**(多个事实fact)和**记忆更新**两步操作, 这样就可以把新的记忆信息永久保存; 2) mem0g, 属于对mem0的一种改进, 用**知识图谱**(knowledge graph)的形式组织记忆知识, 其同样是在每轮交互后进行**记忆信息提取**和**记忆更新**两步操作, 只不过此时记忆信息以三元组(实体1, 实体关系, 实体2)的形式记录, 同样是用llm提取的。
- 在应用memory时, mem0通过对query从记忆库中向量检索, 而mem0g则稍微复杂些, 即用query向量检索记忆中的三元组向量, 也同时从query中提取实体去记忆库中检索得到实体子图。有了记忆后, 就可以扩展到prompt中让llm生成更合适的response。

## 背景

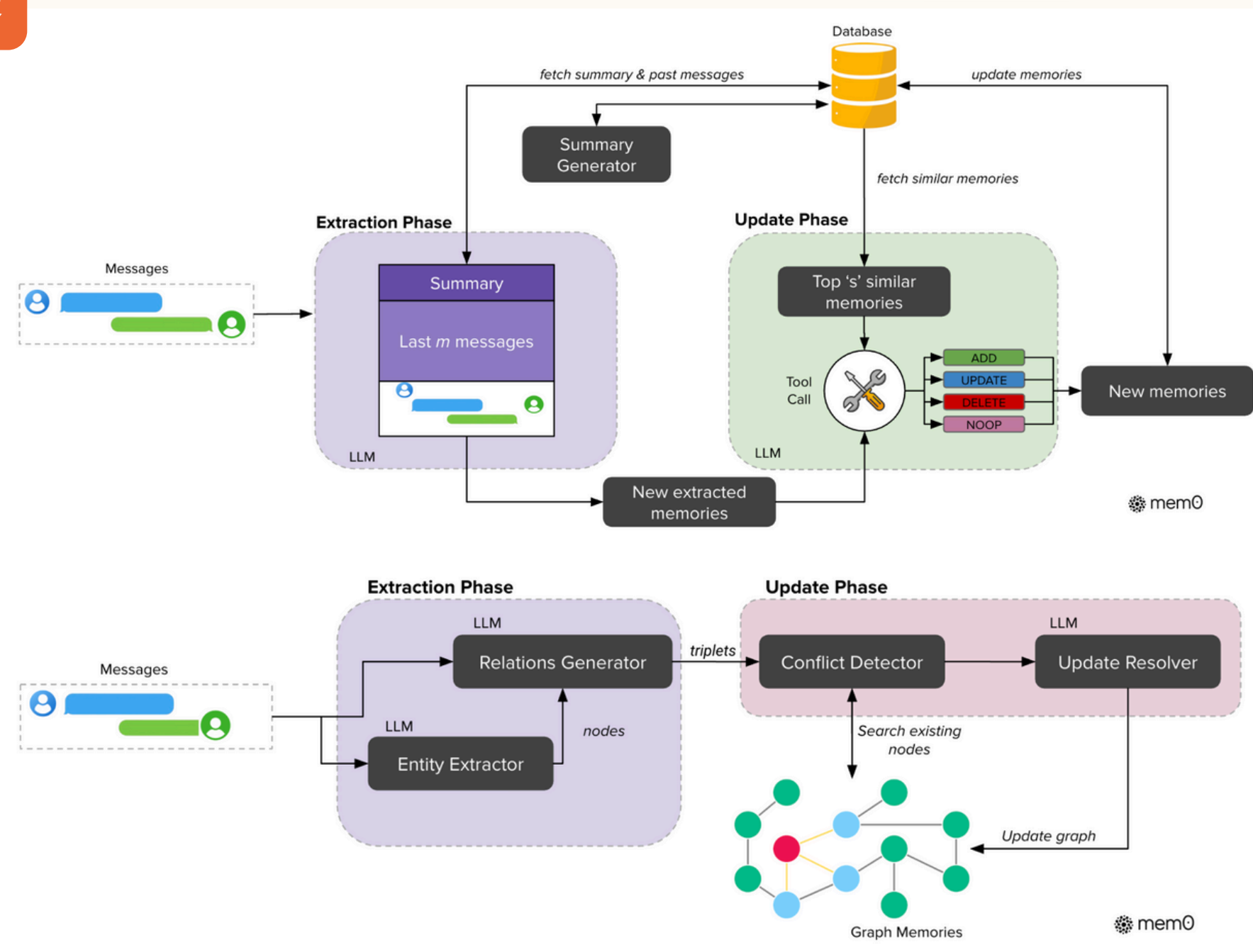
人是有记忆(memory)的, 在漫长的岁月中我们的大脑一直默默进行着信息的提取、整理与更新, 当然也有遗忘, 这些记忆帮助我们跨越时间与情境持续理解世界, 在与其他人交流沟通时帮助理解他人的query以及进行response。

但是, llm却缺乏这种持久的记忆能力, 虽然现在llm的上下文越来越长, 但毕竟是固定大小的, 存在上限, 此外人类记忆还有提取重要信息的功能不仅仅是死记硬背所有的交谈历史。如何为llm或者基于llm的agent设计一套类似人类的memory机制呢? 让我们一起看看mem0给出的方案。

## 实验设置

- 知识图谱数据库Neo4j
- llm一律用GPT-4o-mini

## mem0和mem0g框架



## 思考

主要是出于对ChatGPT中“记忆”机制的好奇, mem0貌似是开源方案中做的很不错的, 读一下技术报告学习学习。