neo4j

AICamp | SINGAPORE

IN PERSON MEETUP

AgenticAi and MCP server

26th June 2025

Links

Neo4j MCP Server
https://neo4j.com/developer/genai-ecosystem/model-context-protocol-mcp/
https://github.com/neo4j-contrib/mcp-neo4j

AuraDB (Neo4j SaaS Graph Database:
https://neo4j.com/product/auradb/https://neo4j.com/product/auradb/

Free Online Courses on Graphs and GenAI:
https://graphacademy.neo4j.com/

GraphRAG Python Neo4j:
https://neo4j.com/docs/neo4j-graphrag-python/current/

LLM Graph Builder:
https://neo4j.com/labs/genai-ecosystem/llm-graph-builder
https://github.com/neo4j-labs/llm-graph-builder

Neoconverse: https://neo4j.com/labs/genai-ecosystem/neoconverse/
https://github.com/neo4j-labs/neoconverse

Xavier Pilas
Senior Solutions Engineer
in Singapore

xavier.pilas@neo4j.com

LinkedIn

Agenda

1. A little bit of LLM history...
2. ... until RAG and GraphRAG...
3. ... to AgenticAI and MCP servers

a LLM is a

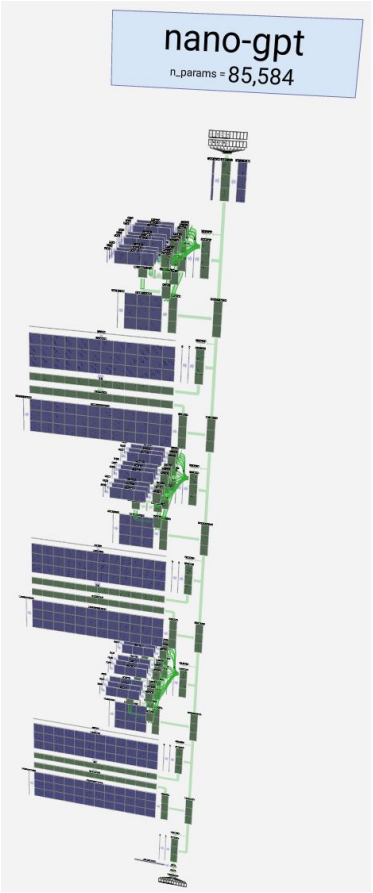
non deterministic function

configured from billions of data

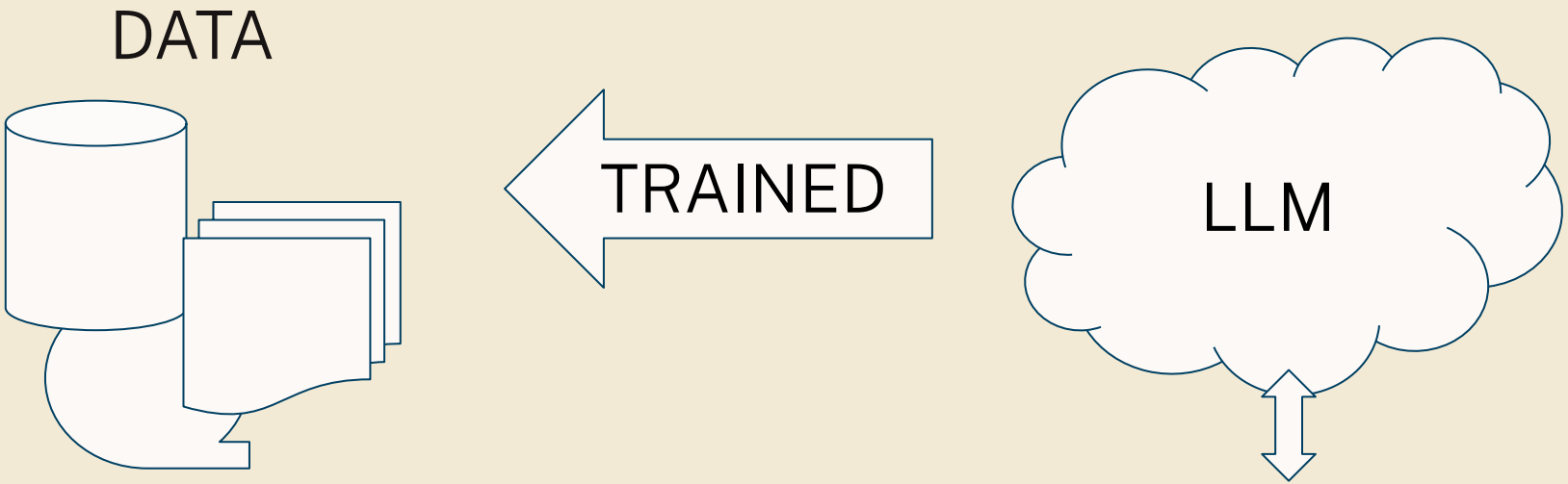
to reproduce **statistically**

answers

https://bbycroft.net/llm



Remember ChatGPT ?



LOTS OF ISSUES

Audit

Bias

Hallucinations

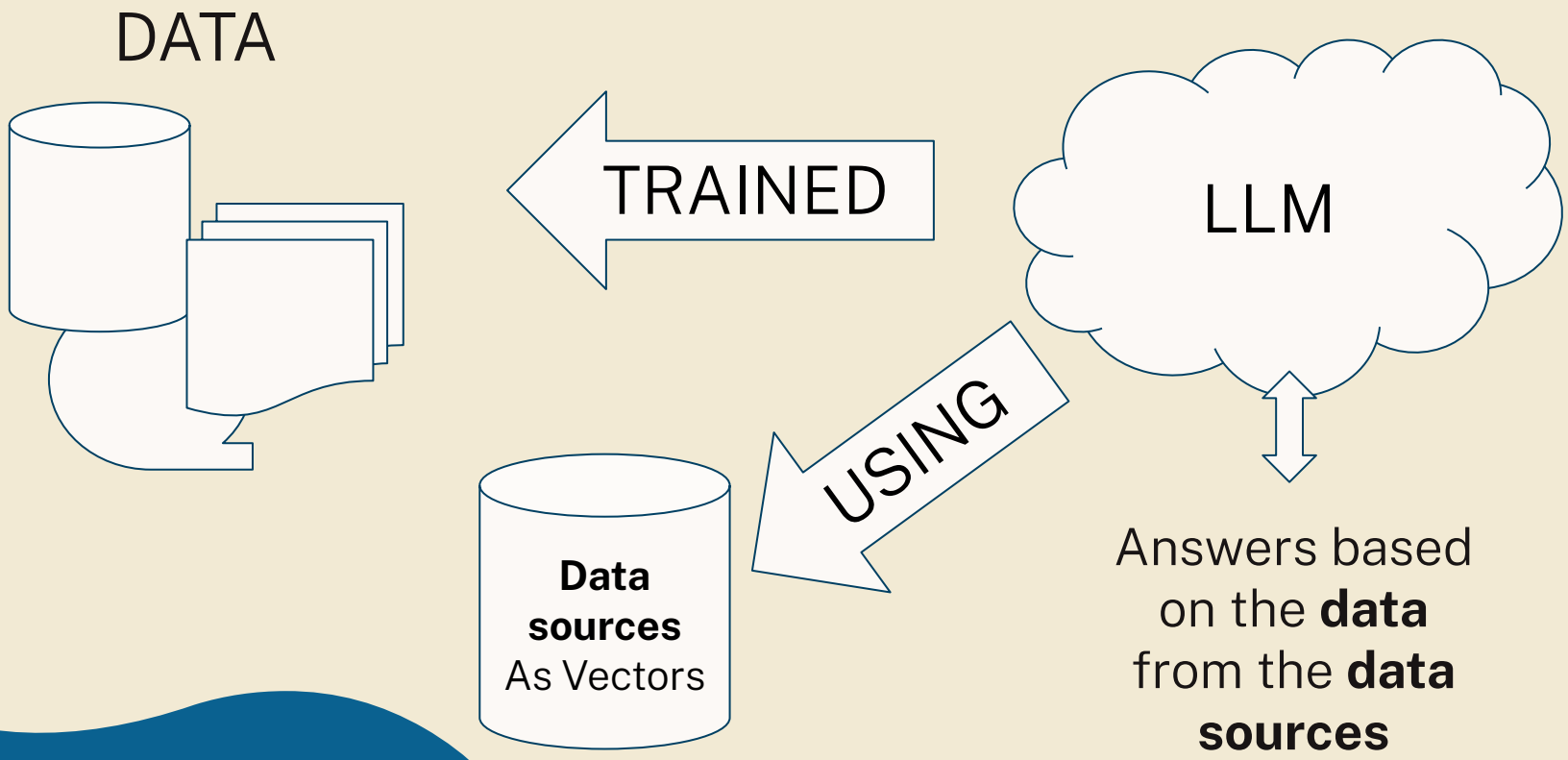
Security

Black box

To use LLMs in my own use case...

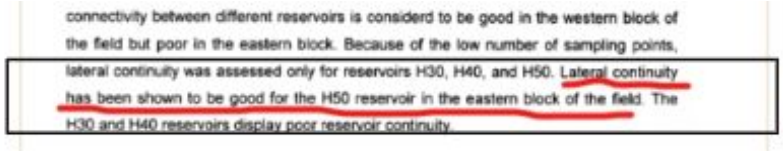
... I want to use my **own** data

Vectors to the rescue, RAG



How does the LLM find the correct data in the data sources ?

1) **Data is stored in chunk which are embedded/vectorized**

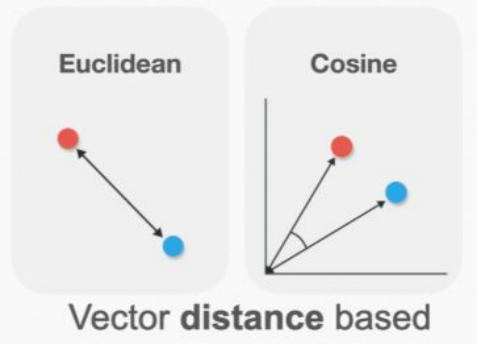


[0.27,-0.09,...]

2) **The LLM transforms the question into a vector:**

Where is Paris ? ***[2.048595, -1.23249, 0.239101, ...]***

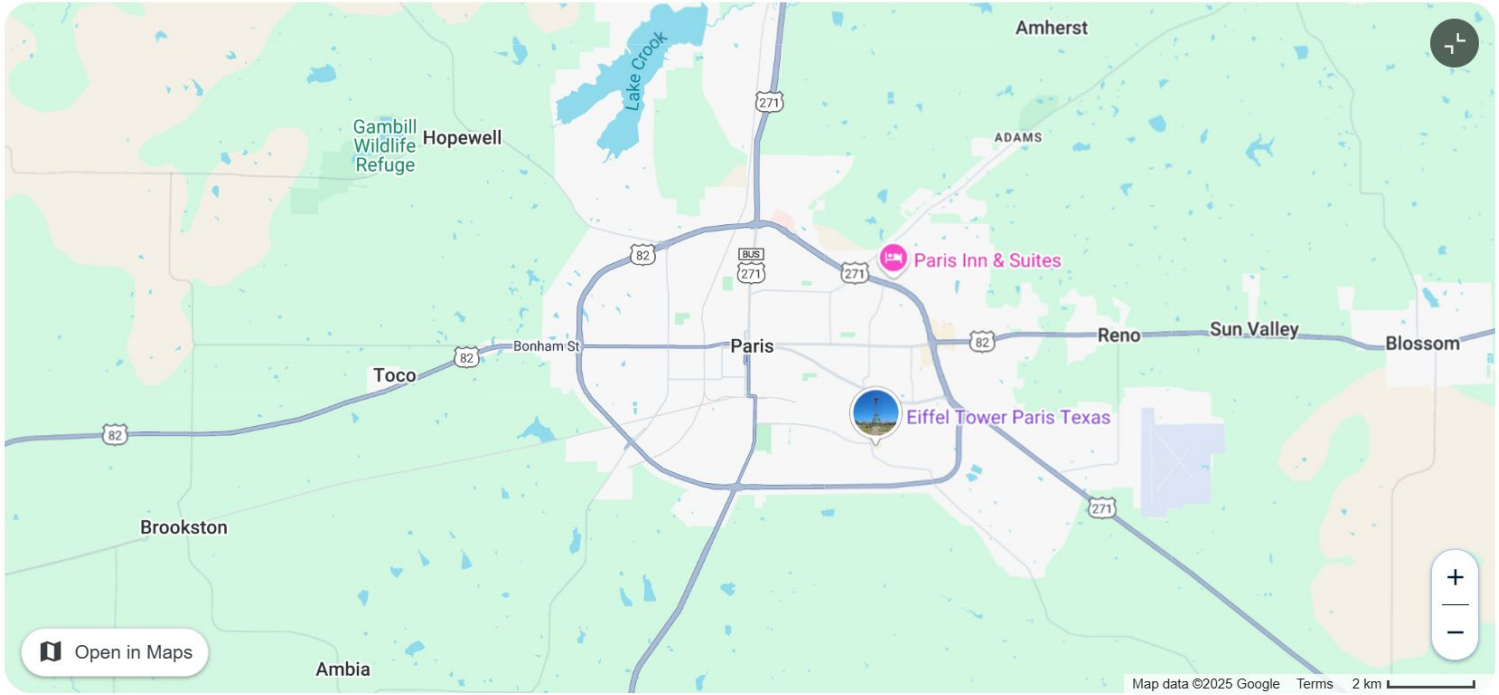
3) **The LLM searches into the data sources the vectors which are the closest**



But it’s not good enough…

Paris

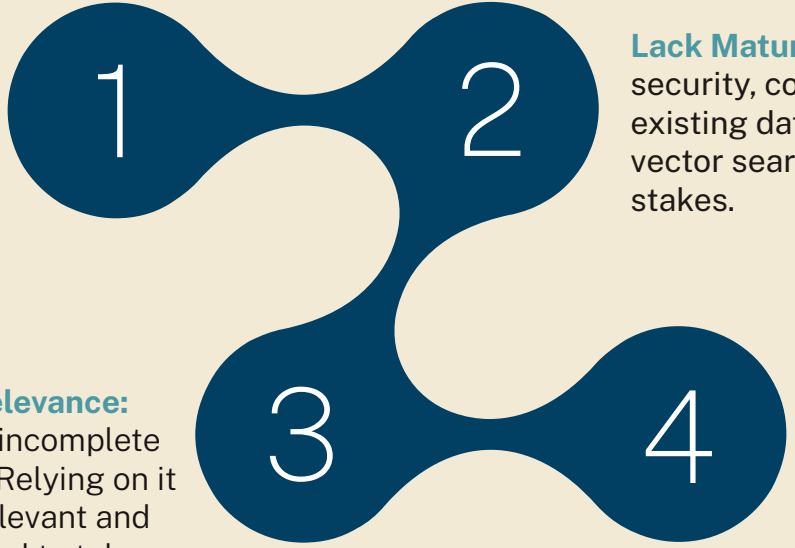
City in Texas



Why RAG With Vector Databases Fall Short

Similarity is insufficient for rich reasoning & explainability

Only leverage a fraction of your data: Beyond simple "metadata", vector databases alone fail to capture relationships from structured data. Lack of structured operations - filter/aggregations/sorting.



Lack Maturity: Don't apply the security, compliance, robustness of existing database systems. Also vector search became DBMS table stakes.

Vector Similarity ≠ Relevance: Vector search uses an incomplete measure of similarity. Relying on it solely can result in irrelevant and duplicative results. Need to take **context** into account.

Lack Explainability: The black-box nature of vectors search lacks transparency and explainability.

The Limitations of Text Embeddings in RAG Applications

Graph is **the** solution

What makes data AI-ready?

Context

Data exposes entities and relationship understanding

Enables AI to grasp meaning and access data across sources

Flexibility

Adapts to new patterns, structures and access logic

Supports diverse query types and evolving data needs

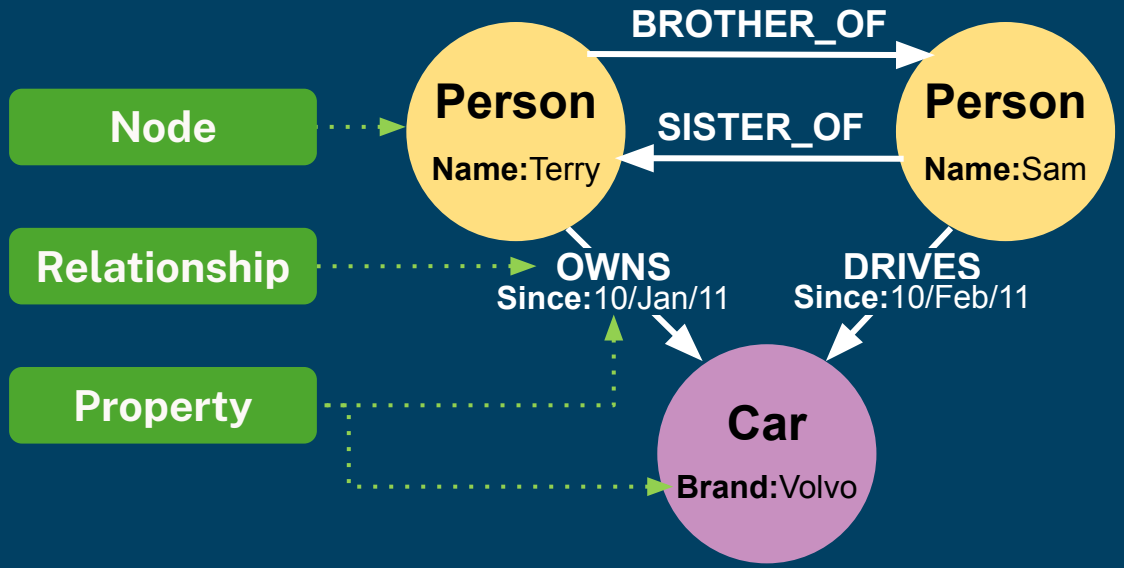
Standardization

Consistent, governed data foundation

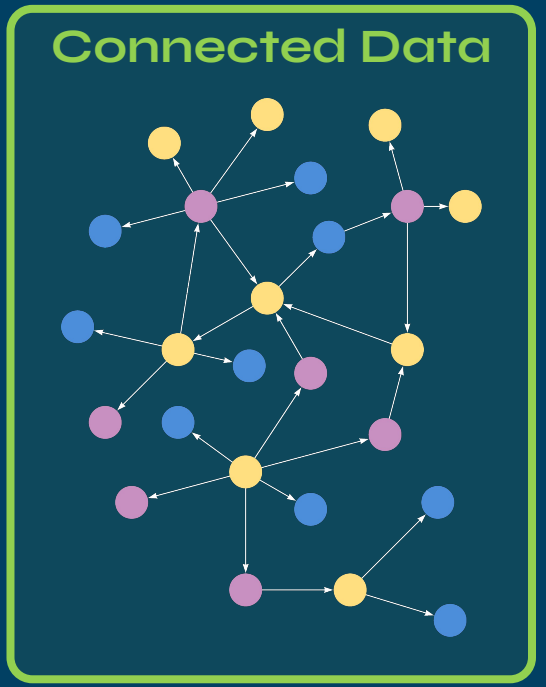
Maintains consistent understanding at scale

Knowledge graphs = AI-ready data

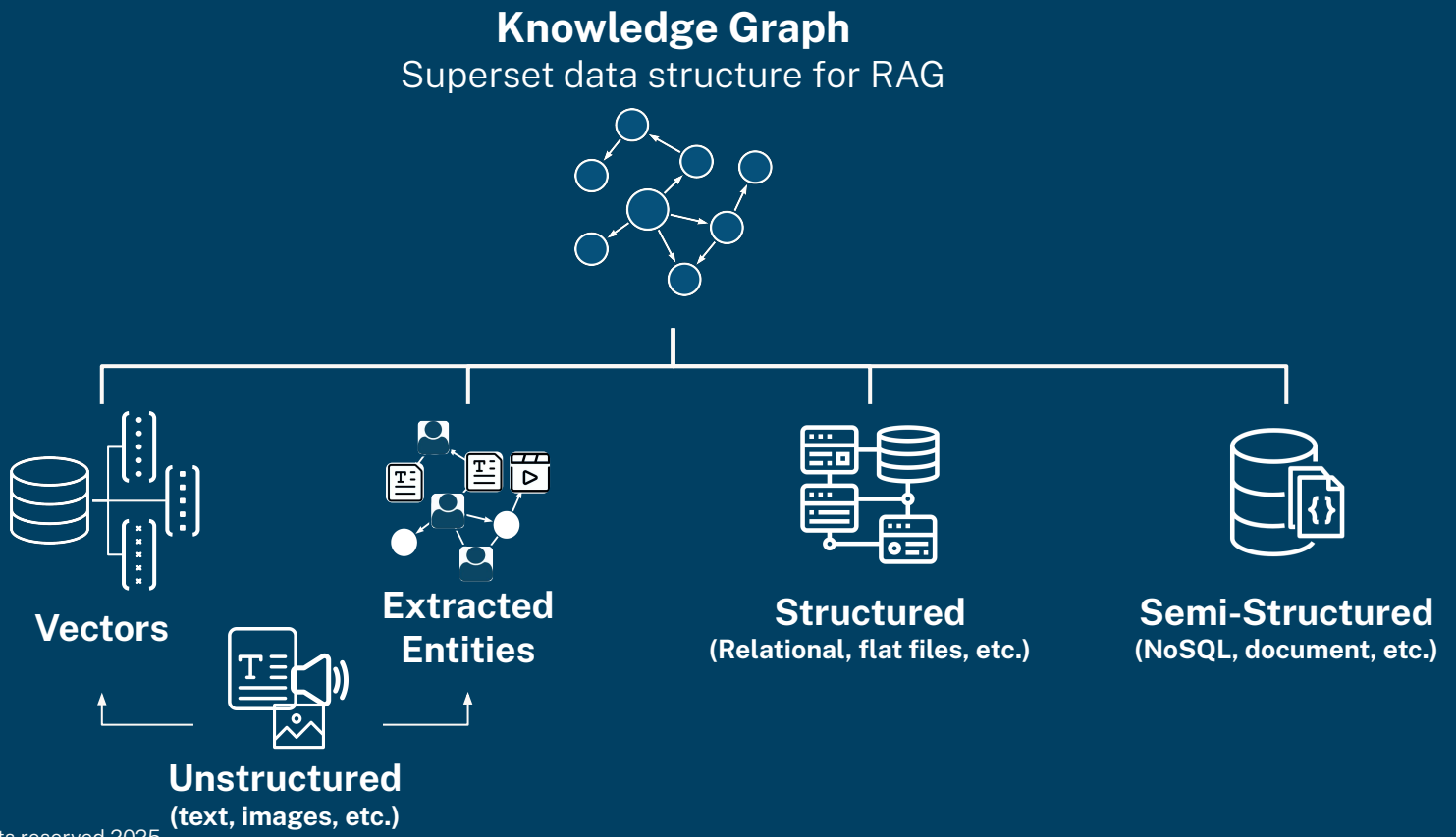
Property Graph Data Model



Connected Data



What to Use as the Data Source for RAG?



GraphRAG: Higher Accuracy Over Traditional RAG

LinkedIn cutdown ticket resolution from 40 → 15 hours using GraphRAG[1]:
1.7x Increase in accuracy over traditional RAG[2]

1.4x accuracy improvement in Lettria Analysis[3]

Up to 2.6x Improvement in RobustQA Benchmark by Writer Knowledge Graph[4]

1) https://www.linkedin.com/feed/update/urn:li:activity:7210725802437443584/
2) https://arxiv.org/pdf/2404.17723
3) https://www.lettria.com/blogpost/vectorrag-vs-graphrag-a-convincing-comparison
4) https://writer.com/engineering/rag-benchmark/

GraphRAG: Higher Accuracy Over Traditional RAG

Zep AI[1]: GraphRAG for AI Agent Memory improves accuracy by 18% and reduce latency by 90%[2]

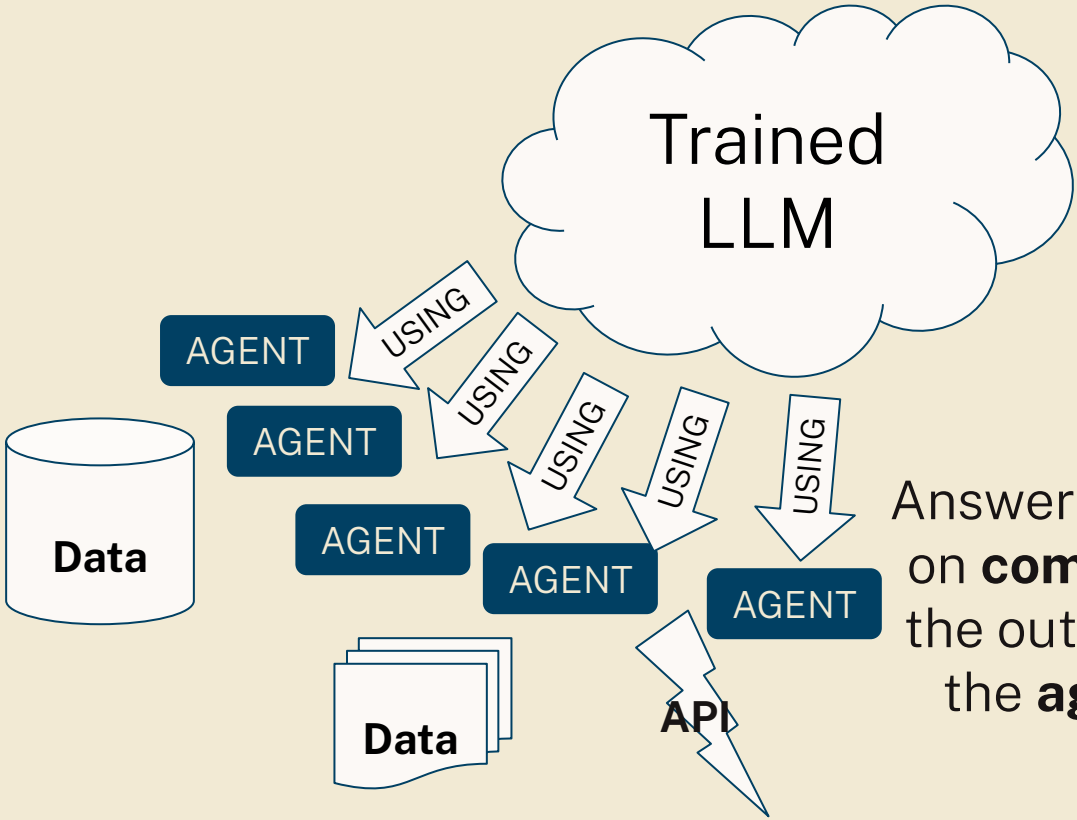
A-Mem: GraphRAG memory systems 2x accuracy for multi-hop reasoning when compared to standard baselines (i.e. ReadAgent, MemGPT, MemoryBank)[3]

LightRAG: GraphRAG outperforms baseline RAG for majority of questions - nearly 85% of instances in legal domain[4]

1) neo4j.com/blog/developer/graphiti-knowledge-graph-memory/
2) https://arxiv.org/pdf/2501.13956
3) https://arxiv.org/pdf/2502.12110
4) https://arxiv.org/pdf/2410.05779

But now we have AgenticAI...

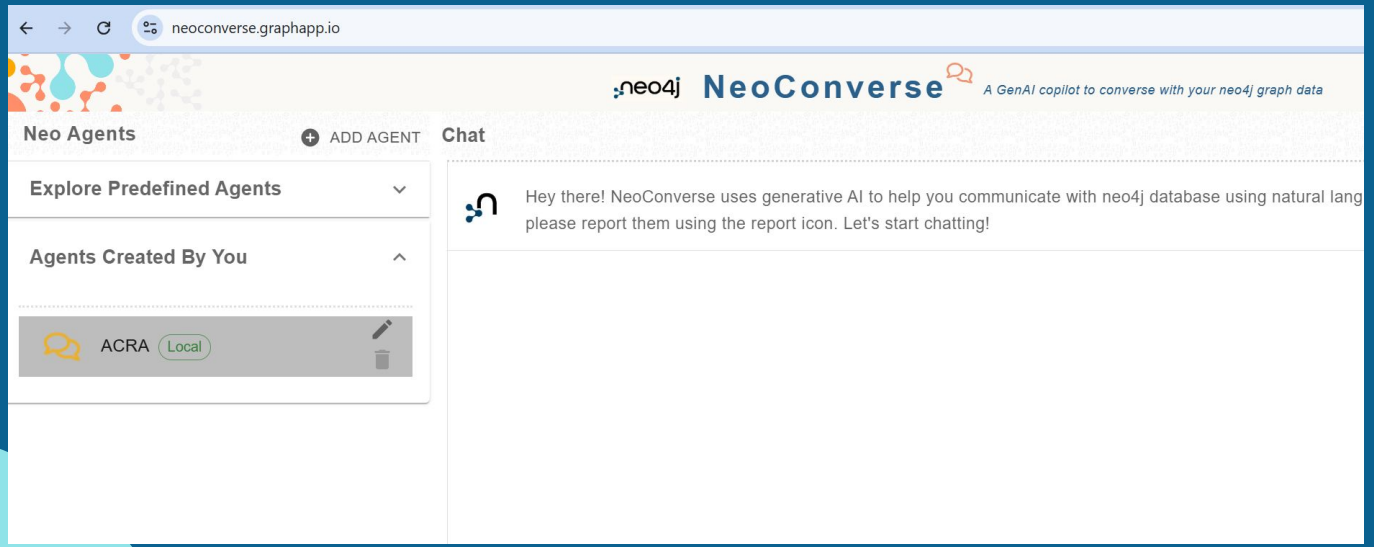
Create your own agents, as many as you need for your use case...



Answers based on **combining** the outputs of the **agents**

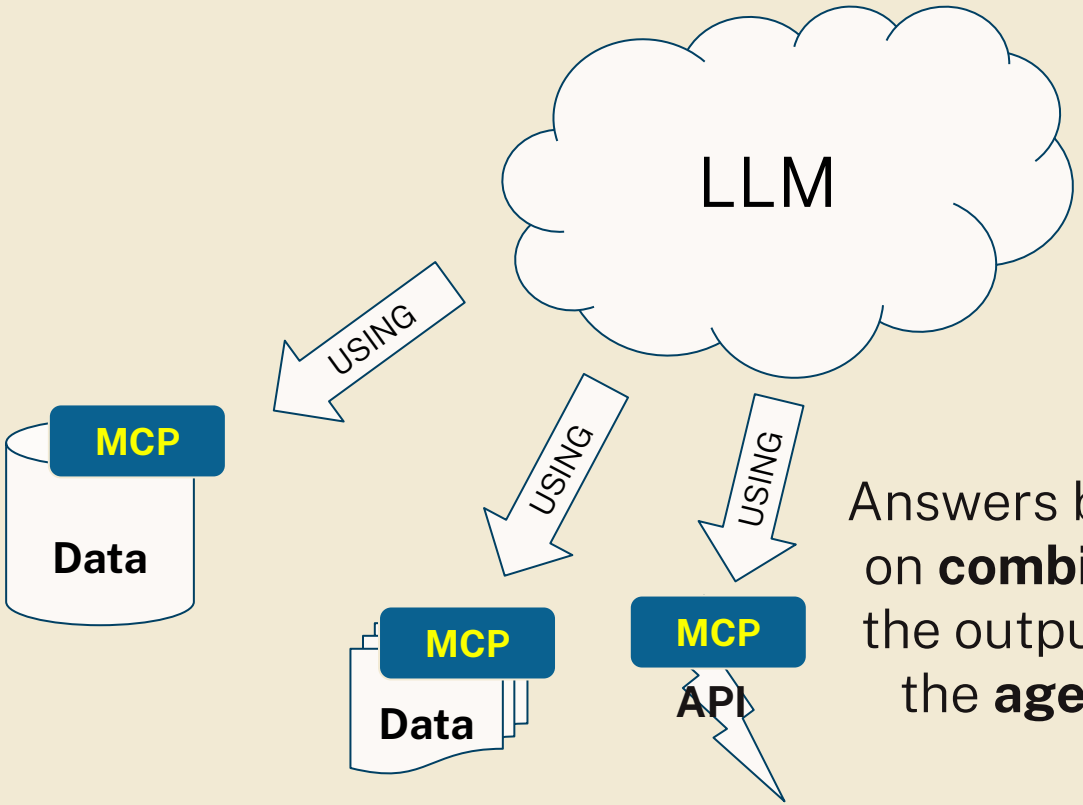
Demo

Neoconverse on ACRA data



What about MCP servers?

<u>One</u> MCP server by source, from the data provider



Answers based on **combining** the outputs of the **agents**

The MCP server will describe the data/service/API in a way the LLM can understand and use it !

Neo4j MCP Server
https://github.com/neo4j-contrib/mcp-neo4j

Demo

Neo4j MCP Server with Gemini still on ACRA data

Questions ?