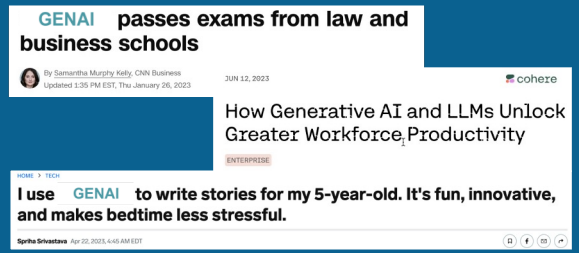

aws
SUMMIT
INDIA

Power Amazon Bedrock applications with Neo4j knowledge graph

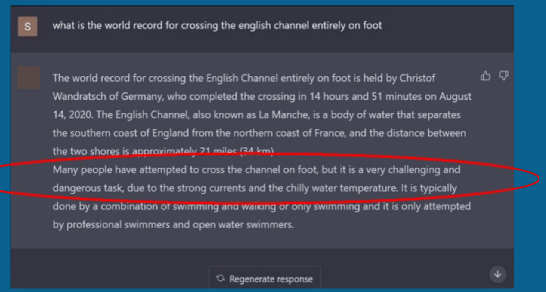
Sumit Shatwara, Sr. Solution Architect, Neo4j India

The State of Generative AI

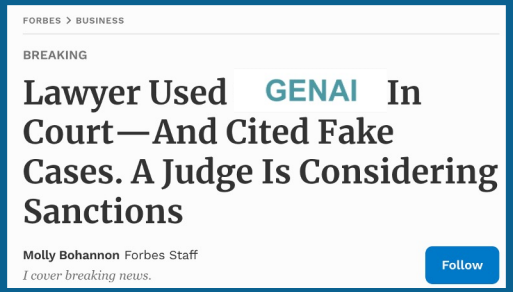
The Good 💪🏽



The Bad 👎🏽



The Ugly 😱



GenAI Alone != Right Outcomes 🤯

Managing AI risk is the biggest barrier to scaling AI initiatives[1]

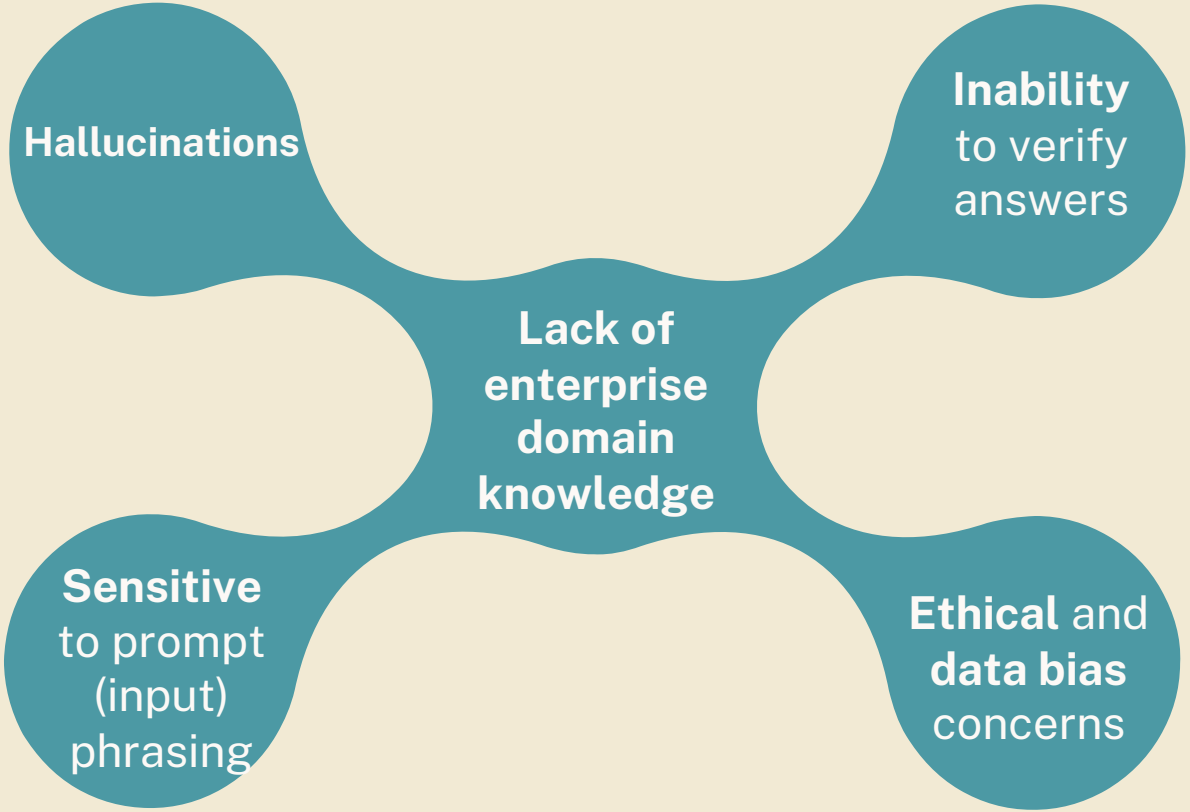
Widespread Hesitancy: Over half of business leaders currently discourage adoption of genAI.[2]

Lack of Explainability: Over 80% of executives worry the non-transparent nature of genAI could result in poor or unlawful decisions.[2]

Risk of Inaccuracy : Inaccuracy and hallucination are two of the most-cited risks of adopting genAI technology at all levels of an organization.[3]

1. Deloitte's State of AI in the Enterprise 2. BCG's Digital Acceleration Index Study 2023 3. McKinsey: The state of AI in 2023

The limitations of LLMs



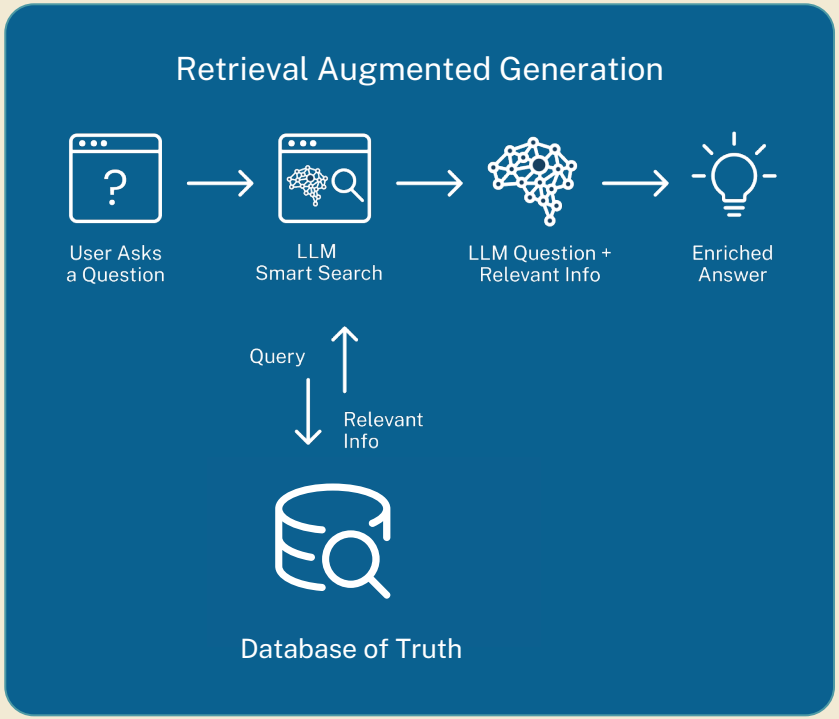
The Big Question

How can organizations use domain-specific knowledge to rapidly build accurate, contextual, and explainable GenAI applications?

Retrieval-Augmented Generation Is Becoming an Industry Standard

RAG augments LLMs by retrieving up-to-date, contextual external data to inform responses:

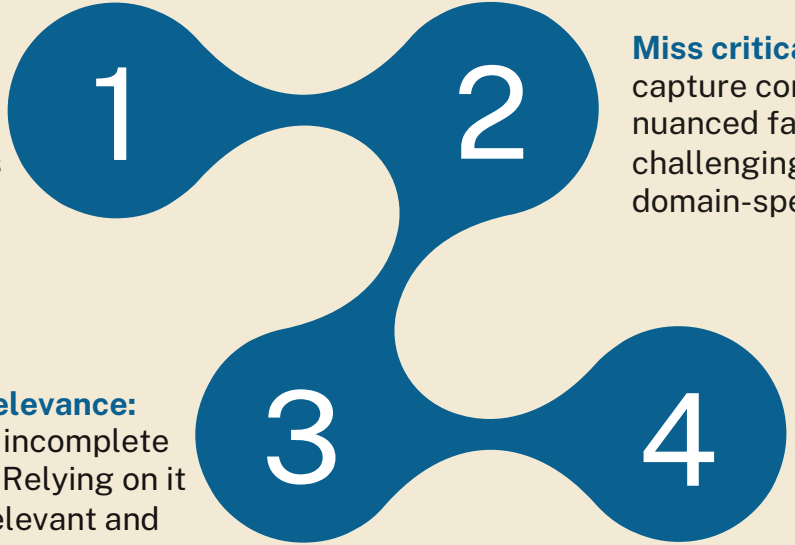
- Reduce hallucinations with verified data
- Provide domain-specific, relevant responses
- Enable traceability back to sources



Why RAG With Vector Databases Fall Short

Similarity is insufficient for rich enterprise reasoning

Only leverage a fraction of your data: Beyond simple “metadata”, vector databases alone fail to capture relationships from structured data



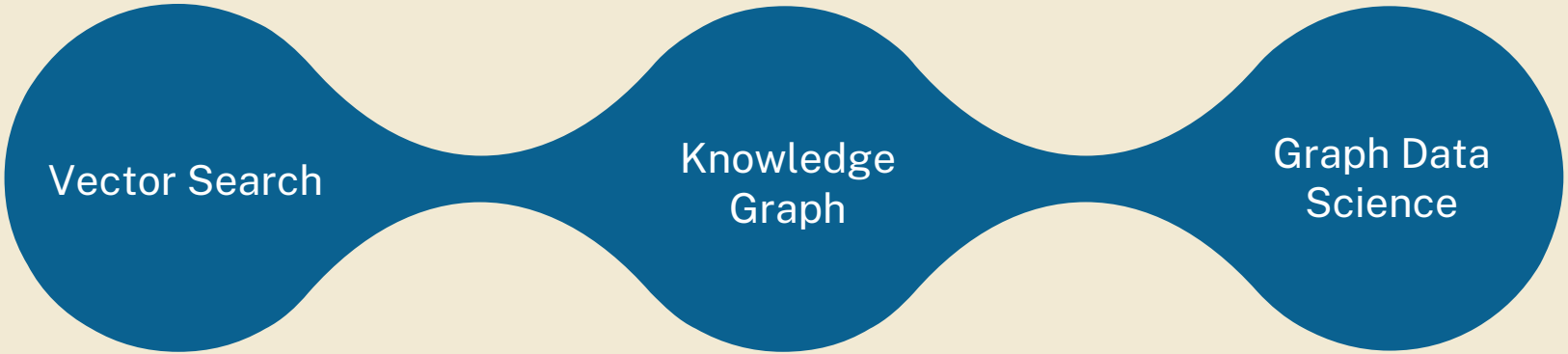
Miss critical context: Struggle to capture connections across nuanced facts, making it challenging to answer multi-step, domain-specific, questions

Vector Similarity ≠ Relevance: Vector search uses an incomplete measure of similarity. Relying on it solely can result in irrelevant and duplicative results

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

Introducing GraphRAG with Neo4j

Unify vector search, knowledge graph and data science capabilities to improve RAG quality and effectiveness



Find similar documents and content

Identify entities associated to content and patterns in connected data

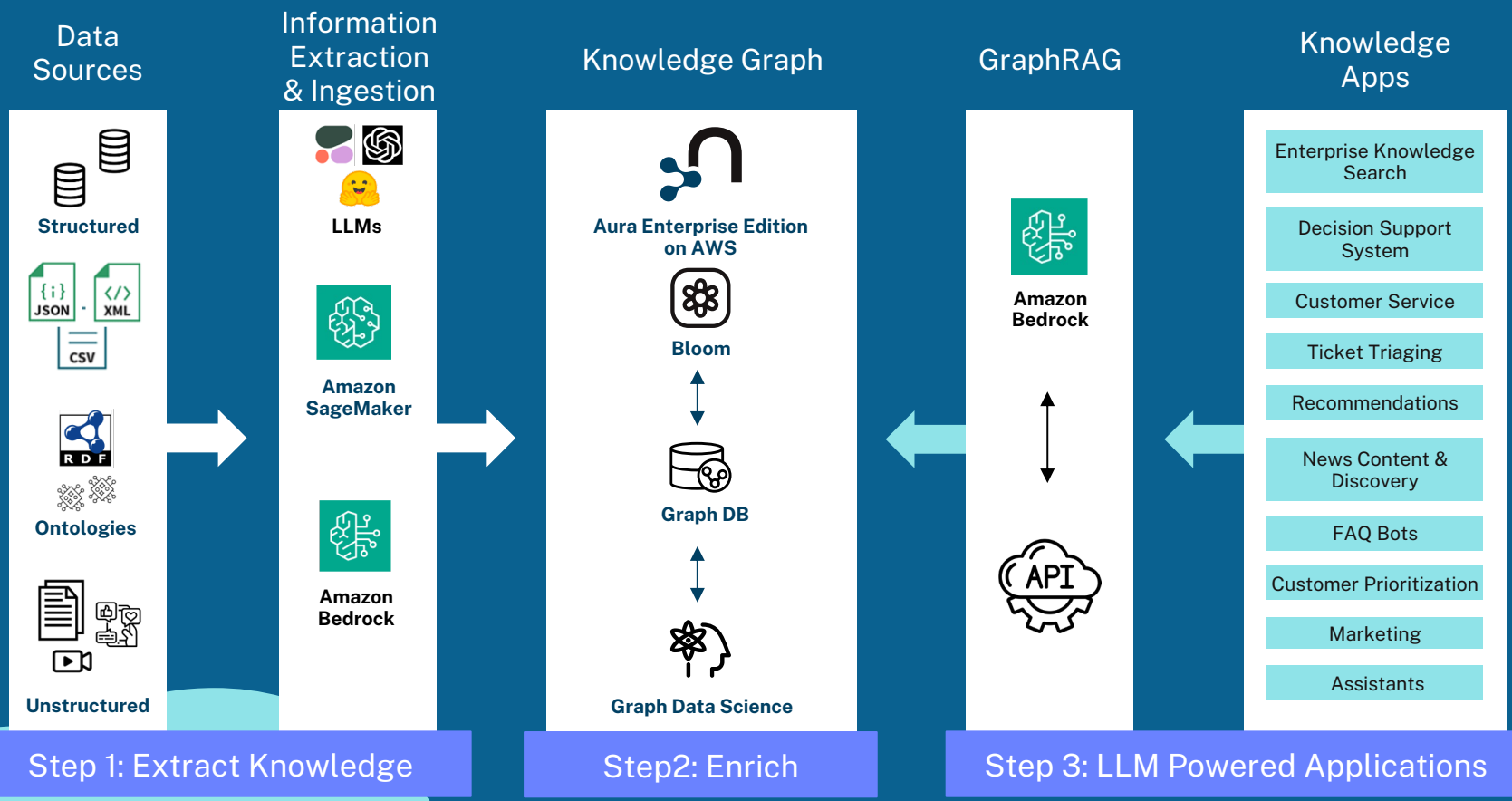
Improve GenAI inferences and insights. Discover new relationships and entities

... to Play a Big Role in the Future of GenAI

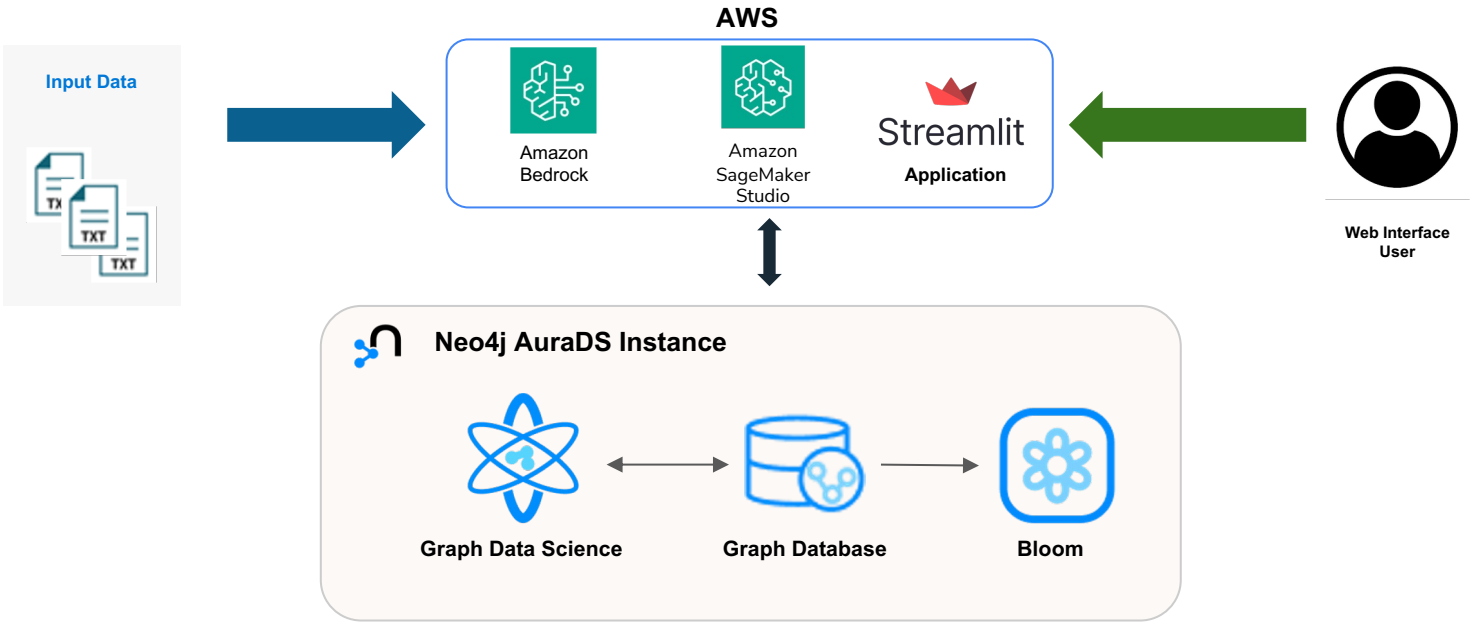
> By 2025, 50% of generative AI initiatives will have improved reliability and transparency by combining deep learning foundation models with knowledge graphs or other composite AI elements.”

Source: Gartner®, Technological Implications of Generative AI (August 2023)

Neo4j & GenAI Reference Architecture on AWS

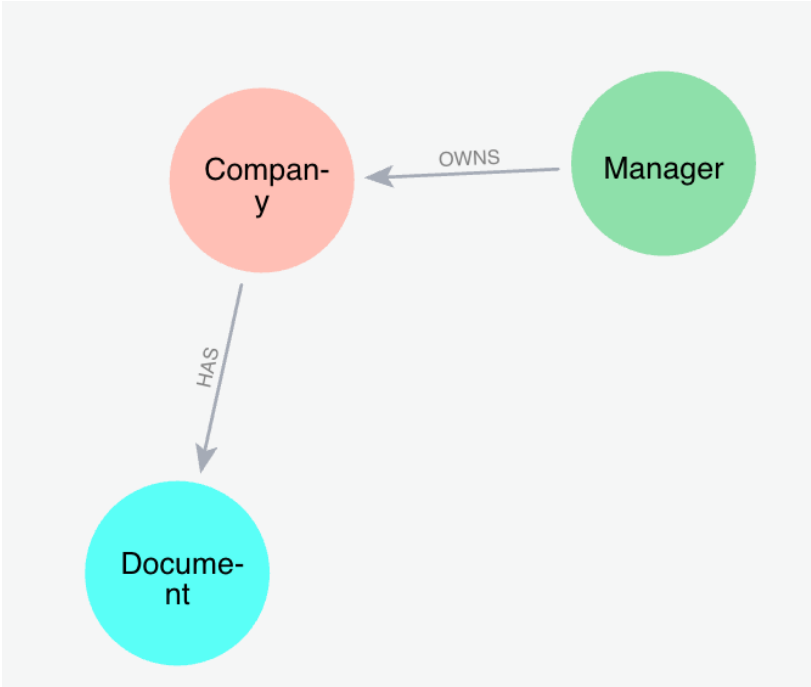


Demo Architecture



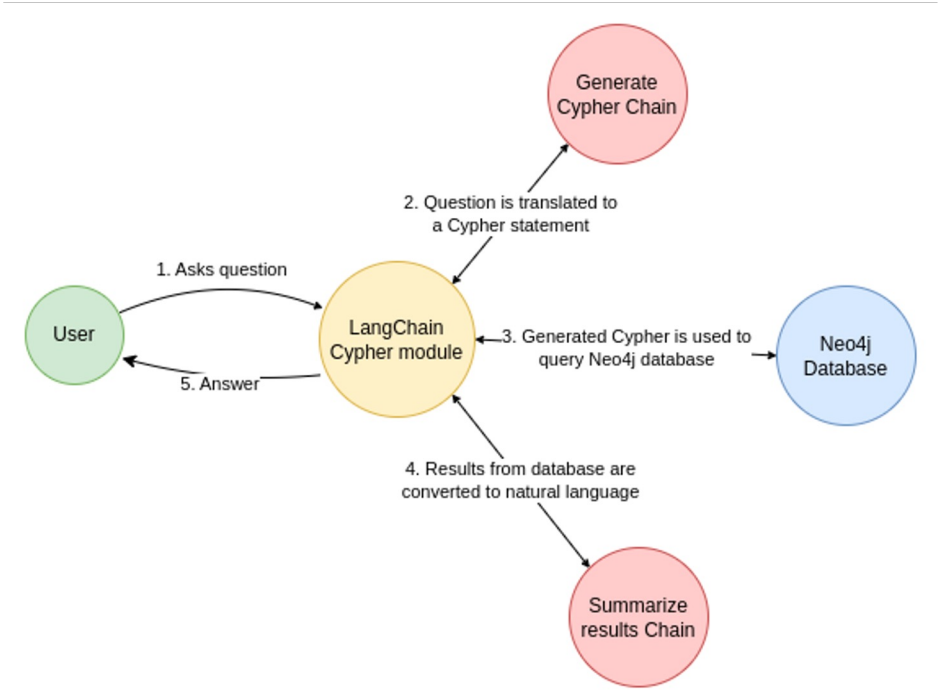
Parsing

- Zero-shot with a simple prompt with the LLM
- Extract SEC EDGAR filing information in accordance with a Neo4j data model



Chatbot

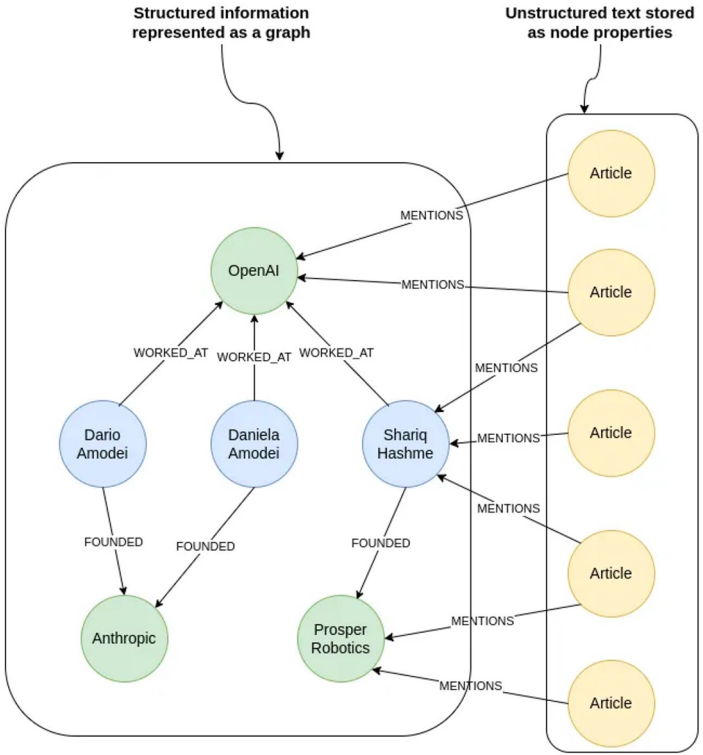
- Translates English to Cypher
- Consumption using LLM with few shot prompting
- Data augmentation from Neo4j response



Semantic Search

If your focus is analyzing documents on a file system, then vector indexing and search on text embeddings may be sufficient.

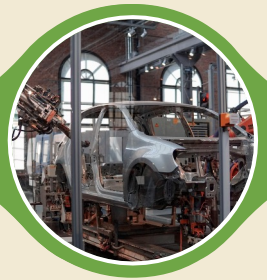
If you need to retrieve and make inferences about people, places, and things connected to those documents, knowledge graphs can help.



Real-World GenAI Breakthroughs Powered by Neo4j

Leveraging AI for customized content at scale

Integrating AI with knowledge graphs for smarter supply chain management



Merging structured and unstructured data for efficient operations

Converting vast amounts of unstructured data into actionable knowledge

Flexible Cloud Deployment Models on AWS

Graph-as-a-Service

Cloud-native
Self-service deployment
No access to underlying
infrastructure and systems

Cloud Managed Services

Fully customizable deployment
model and service levels
Operate In own data centers
or Virtual Private Cloud

Self-Hosted

Bring your own license
Full control of your environment
Run in any cloud, in your account

Getting Started with Neo4j Aura on AWS

GenAI Bundles

Fully-Managed Bundles*:

Starter: (24 vCPU, 128GB)
Regular: (48 vCPU, 256GB)
Large: (82 vCPU, 384GB)

All bundles include:

- Graph Database
- 70+ AI & ML Algorithms
- 5 user licenses for Bloom for visualization
- Connectors (Data Warehouse, BI, Spark, etc.)

AWS Partnership

- Pay with Cloud Credits
- Virtual Private Clouds Available
- Amazon Bedrock

Training Offers

- Cloud Partner Workshops
- Hands-on Labs
- Graph Academy

*Negotiated and Self-Managed bundles also available

Thank you