neo4j graphsummit

Knowledge Graphs and Generative AI

Dr. Katie Roberts

Data Science Solutions Architect

Neo4j

Agenda

Grounding LLMs with Knowledge Graphs

- Knowledge Graphs
- Graph Data Science
- Semantic Search

Neo4j Grounded Chatbot Demo

Katie Roberts

Data Science Solution Architect

LLMs Give Us an Amazing Opportunity to:

1. Automate data retrieval tasks
2. Improve customer service experiences
3. Expedite reading, understanding, & summarizing
4. Generate content & code

Example...

Q: What was the impact of Hurricane Lee?

- A: Hurricane Lee was a long-lasting tropical cyclone that formed in the Atlantic during the 2017 hurricane season.....

Hurricane Lee

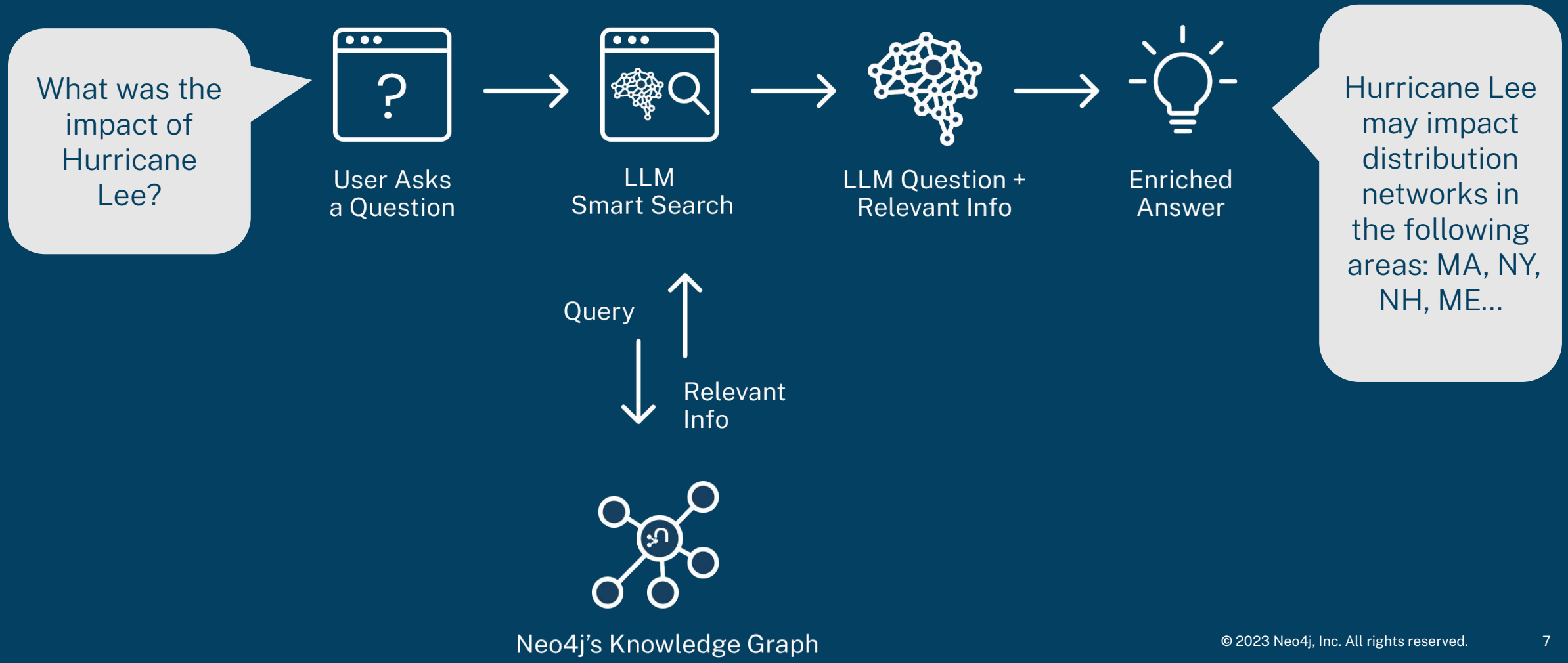
Sep 5, 2023 – Sep 19, 2023

But There Are Challenges...

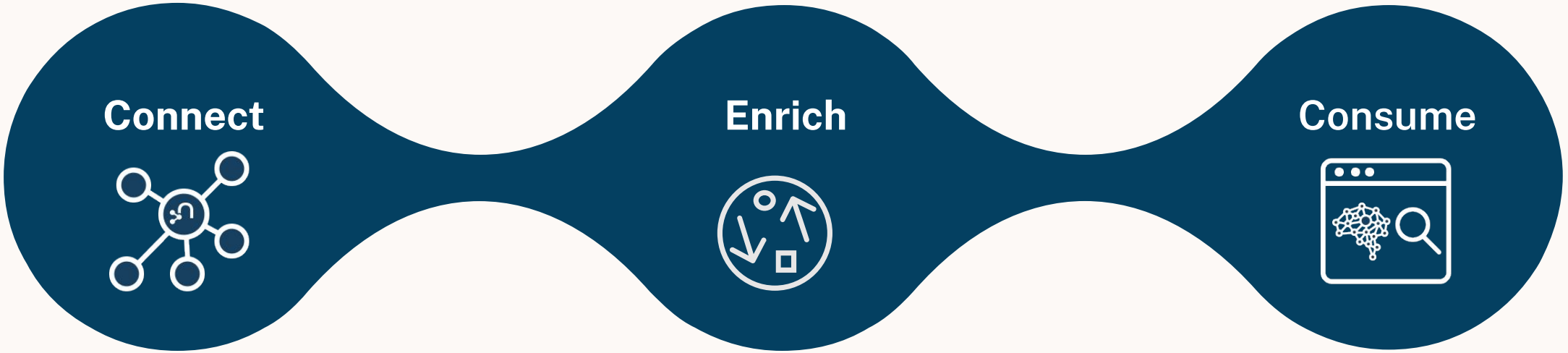
1. Knowledge cut-off
2. Reasonable answers, not always accurate
3. Can inherit bias through training data
4. Lack of enterprise domain knowledge
5. Inability to verify or attribute sources

How can you take advantage of this massive opportunity while overcoming these challenges?

Use a Knowledge Graph to Ground LLMs



Grounding with Neo4j Knowledge Graphs



Context rich, connected view of your data that enables easier decision making

Enhance your data with **graph data science**, text embeddings, and additional derived context

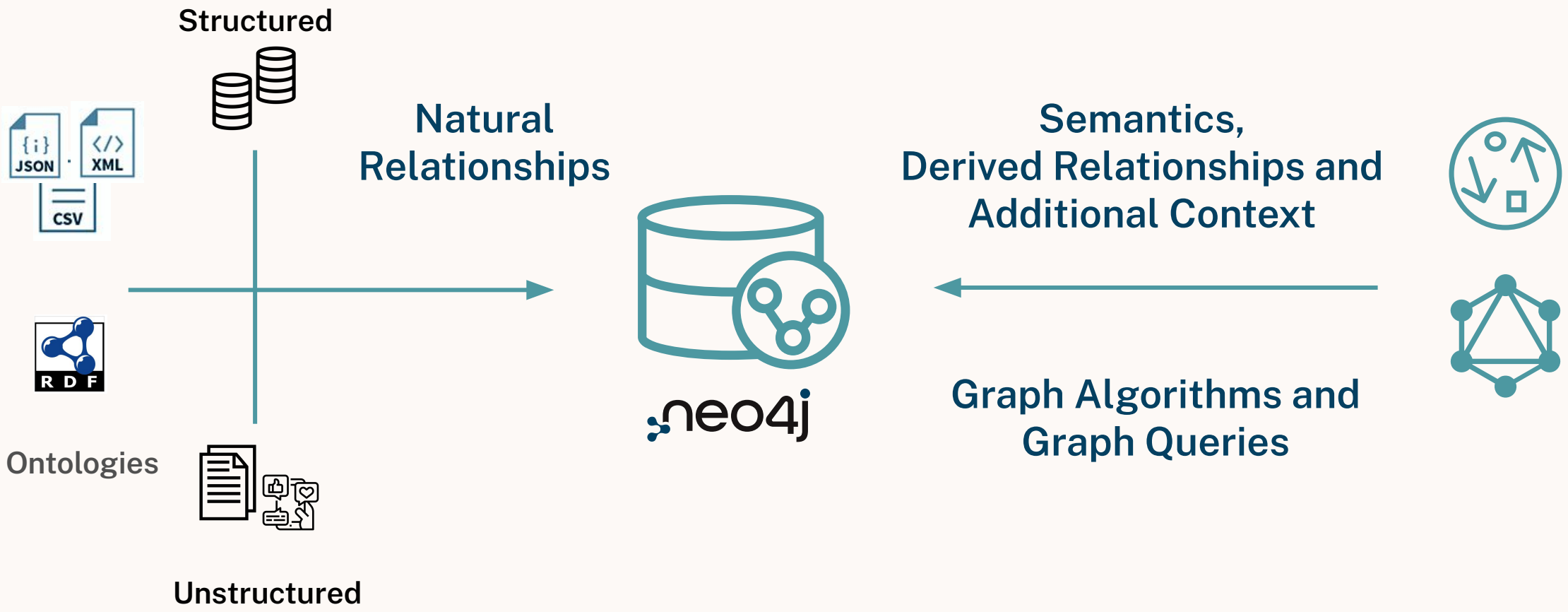
Ground responses with information and context in the graph

I**mprove search relevance** combining vector search and graph traversals

Connected Data

With Knowledge Graphs

Knowledge Graphs



What can you do with a knowledge graph?

neo4j graphsummit

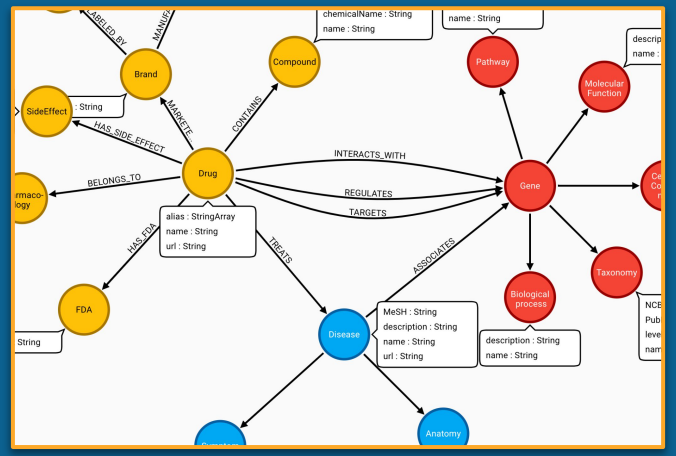
Finance

How many flagged accounts are in the applicant's network **4+ hops out?**

How many **login / account variables in common**?

Add these metrics to your approval process

Life Sciences

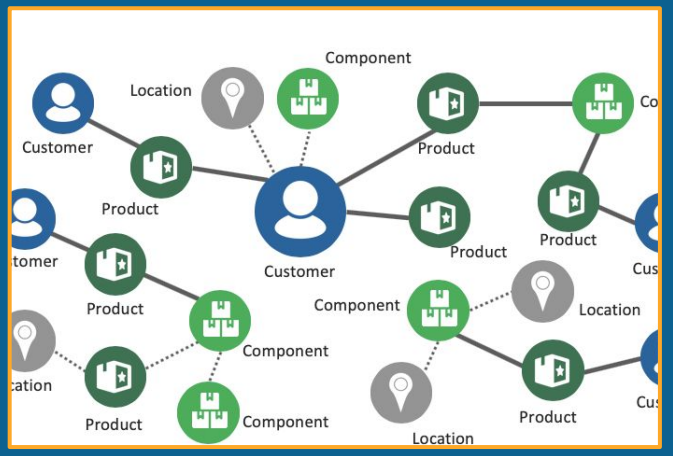


What **completes the connections** from genes to diseases to targets?

What genes can be reached **4+ hops out** from a known drug target?

What **mechanisms in common** are there between two drugs?

Marketing & Recommendations



Collaborative filtering: users who bought X, also bought Y

What items make you more likely to buy additional items **in subsequent transactions**?

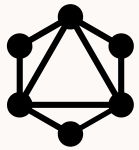
Traverse hierarchies - what items are similar **4+ hops out?**

Enrich

With Cypher Patterns, Graph Data Science & Text Embeddings

Graph Enrichment

Queries



Human-crafted query, human-readable result

```
MATCH (p1:Person)-[:ENEMY]->(:Person)<-[:ENEMY]-(p2:PERSON)
MERGE (p1)-[:FRIEND]->(p2)
```

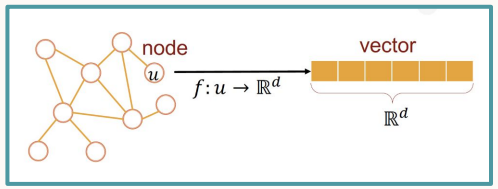
Algorithms

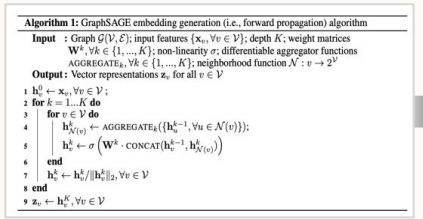
Predefined formula, human-readable result

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$$

```
PageRank(Emil)    = 13.25
PageRank(Amy)     =  4.83
PageRank(Alicia)  =  4.75
```

Embeddings





Algorithm 1: GraphSAGE embedding generation (i.e., forward propagation) algorithm

Input : Graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$; input features $\{\mathbf{x}_v, \forall v \in \mathcal{V}\}$; depth K; weight matrices $\mathbf{W}^k, \forall k \in \{1, ..., K\}$; non-linearity σ; differentiable aggregator functions $\text{AGGREGATE}_k, \forall k \in \{1, ..., K\}$; neighborhood function $\mathcal{N} : v \rightarrow 2^{\mathcal{V}}$

Output: Vector representations $\mathbf{z}_v$ for all $v \in \mathcal{V}$

$\mathbf{h}_v^0 \leftarrow \mathbf{x}_v, \forall v \in \mathcal{V}$;
for $k = 1...K$ **do**
  for $v \in \mathcal{V}$ **do**
    $\mathbf{h}_{\mathcal{N}(v)}^k \leftarrow \text{AGGREGATE}_k(\{\mathbf{h}_u^{k-1}, \forall u \in \mathcal{N}(v)\})$;
    $\mathbf{h}_v^k \leftarrow \sigma\left(\mathbf{W}^k \cdot \text{CONCAT}(\mathbf{h}_v^{k-1}, \mathbf{h}_{\mathcal{N}(v)}^k)\right)$
  end
  $\mathbf{h}_v^k \leftarrow \mathbf{h}_v^k / \|\mathbf{h}_v^k\|_2, \forall v \in \mathcal{V}$
end
$\mathbf{z}_v \leftarrow \mathbf{h}_v^K, \forall v \in \mathcal{V}$

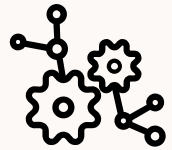
AI-learned formula, machine-readable result

```
Node2Vec(Emil)   =[5.4 5.1 2.4 4.5 3.1]
Node2Vec(Amy)    =[2.8 1.8 7.2 0.9 3.0]
Node2Vec(Katie)  =[1.4 5.2 4.4 3.9 3.2]
```

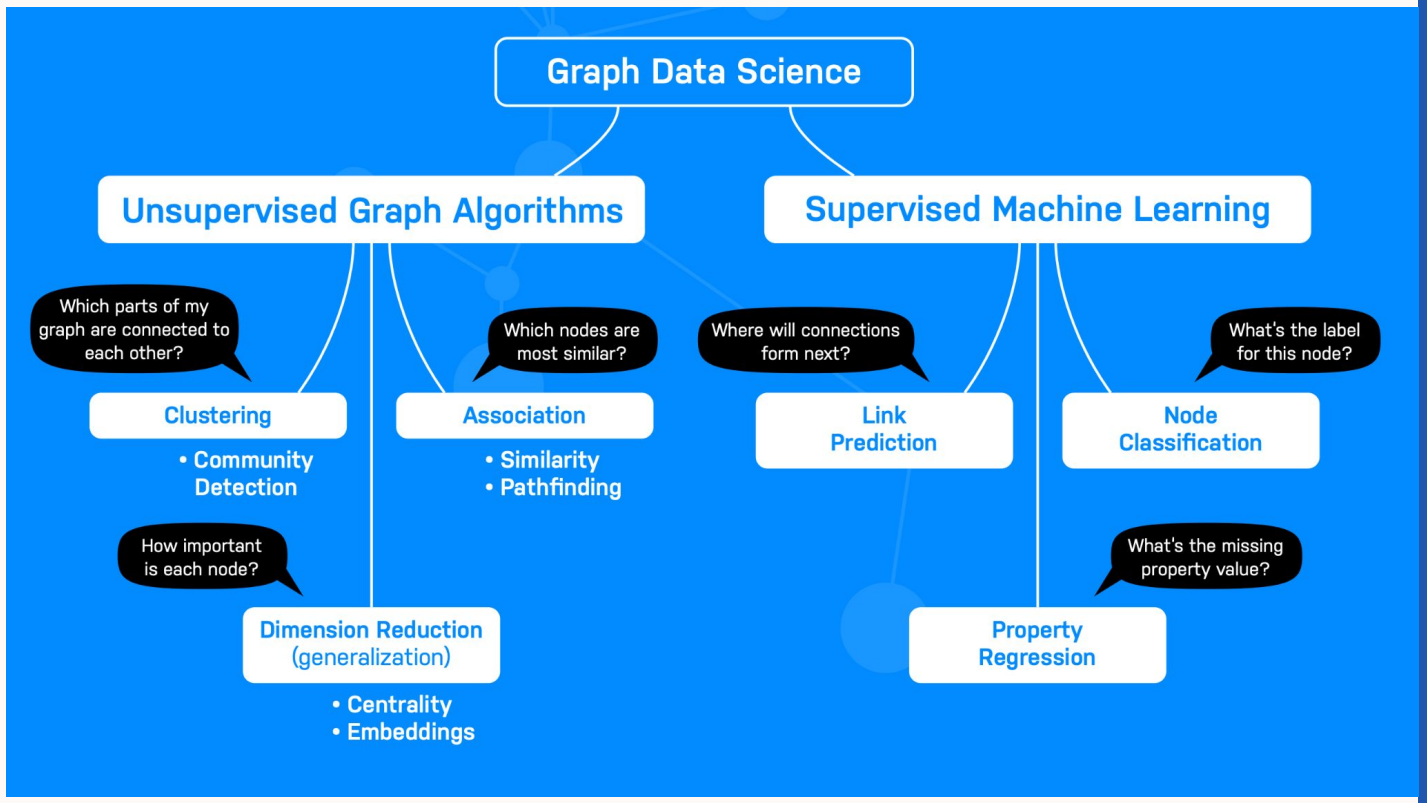
Machine Learning Workflows

Train ML models based on results

$$f(x)$$



What Are Graph Algorithms?



Insights From Graph Algorithms

Centrality Outliers, Influencers, Vulnerabilities,..

Pathfinding Shortest Path, Optimal path, Route Optimization,...

Community Detection Recommendations, Homophily, Outliers,..

Similarity Recommendations, What-if Analysis, Disambiguation,..

Embeddings Dimensionality Reduction, Representation Learning, ..

Link Prediction Link prediction, Recommendations, Next-Best Action,..

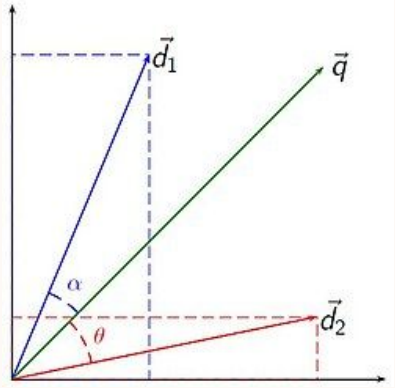
Text Embedding Vectors for Semantic Search

Given a question, find the most relevant documents based on a similarity metric (such as Cosine Similarity) between vector of the question and vectors of contents.

Moving from keyword search to similarity (semantic) search.

Q: what is a text embedding?

embedding	[-0.019200351,0.0035306285,0.014364655,-0.009949144,0.0011581815,0.017021084,-0.01998375,-0.018986698,-0.033643346,-0.02841595,0.014685135,0.014193732,... Show all]

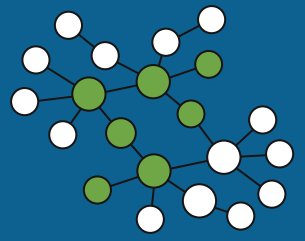


Top K by similarity

abstractId	similarity
456	0.923445
22	0.892114
...	...

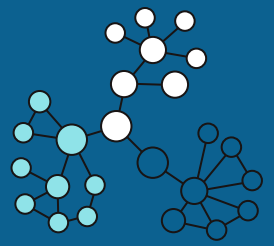
Graph Data Science Journey

Knowledge Graphs



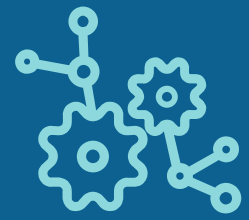
Find the patterns you're looking for in connected data

Graph Algorithms



Identify associations, anomalies, and trends using unsupervised machine learning

Graph Native ML



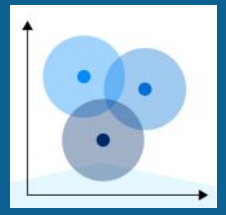
Learn features in your graph that you don't even know are important yet

Consume

With LLMs and Semantic Search

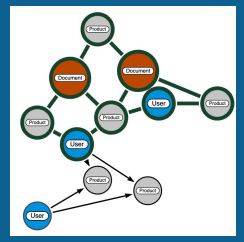
Semantic Search Journey

Vector Similarity Search



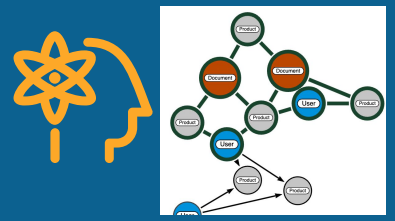
Find relevant documents and content for user queries

Graph Traversals & Pattern Matching



Find people, places, and things associated to content. Identify patterns in connected data.

Knowledge Graph Inference & ML

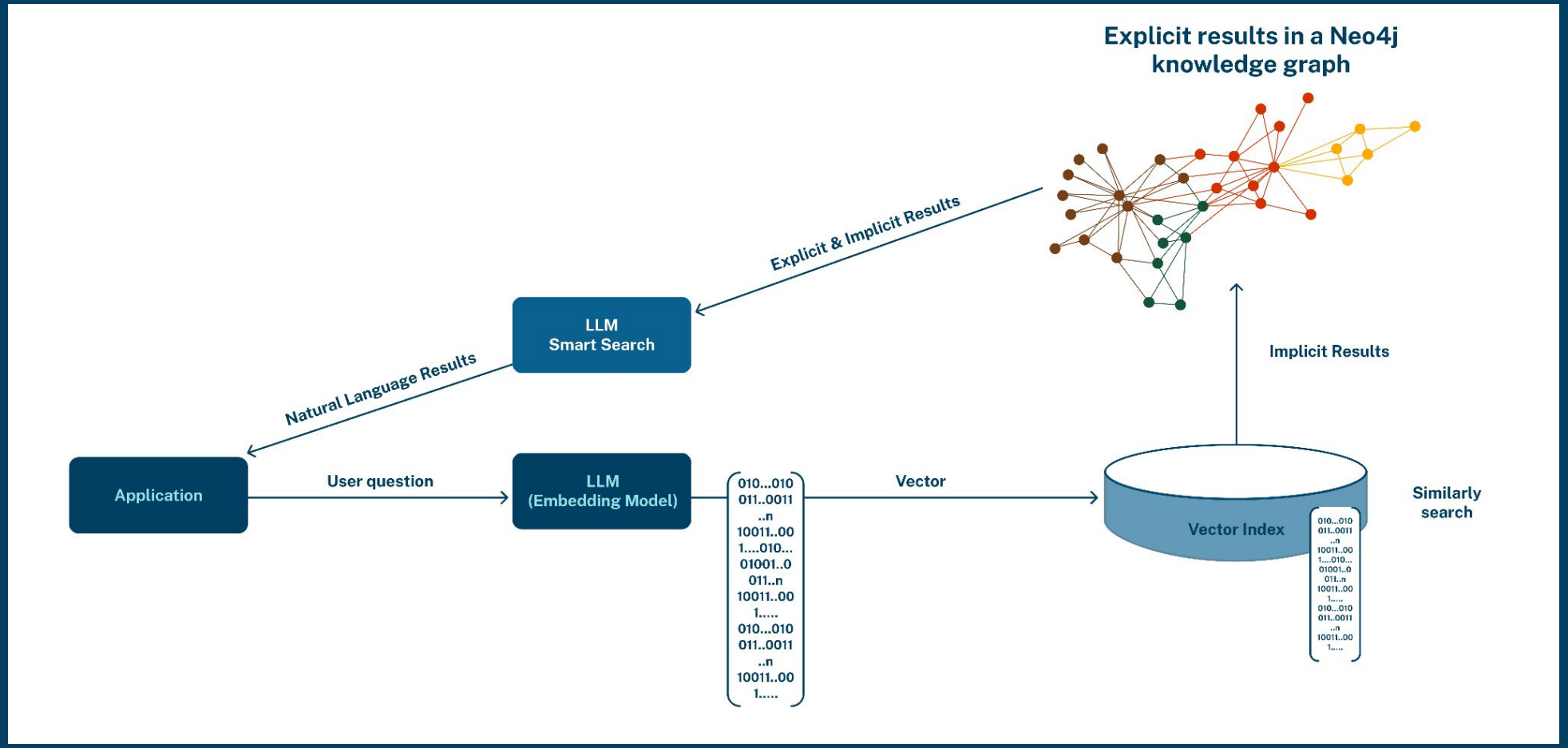


Further improve search relevance and insights by enhancing your Knowledge Graph.
Use graph algorithms and ML to discover new relationships, entities, and groups.

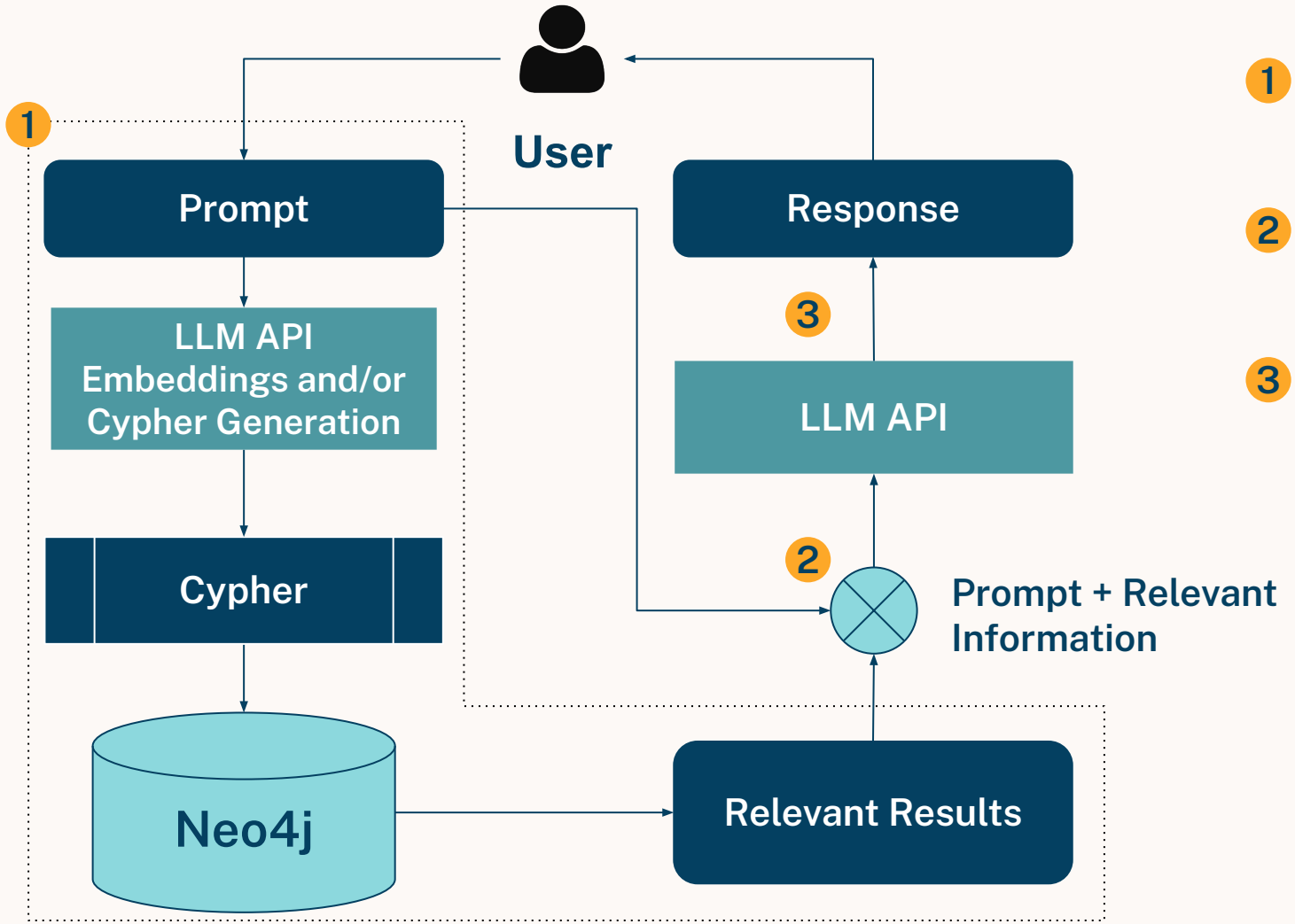
Vector Database | **Graph Database** | **Graph Data Science**

Neo4j

Knowledge Retrieval with Neo4j



RAG (Retrieval Augmented Generation) Pattern with Neo4j



1. Retrieve relevant results from Neo4j using LLM to generate embeddings and/or Cypher
2. Combine relevant results with prompt
3. Instruct LLM to only use the relevant results to generate response

Improved ACCURACY and RELEVANCE of responses

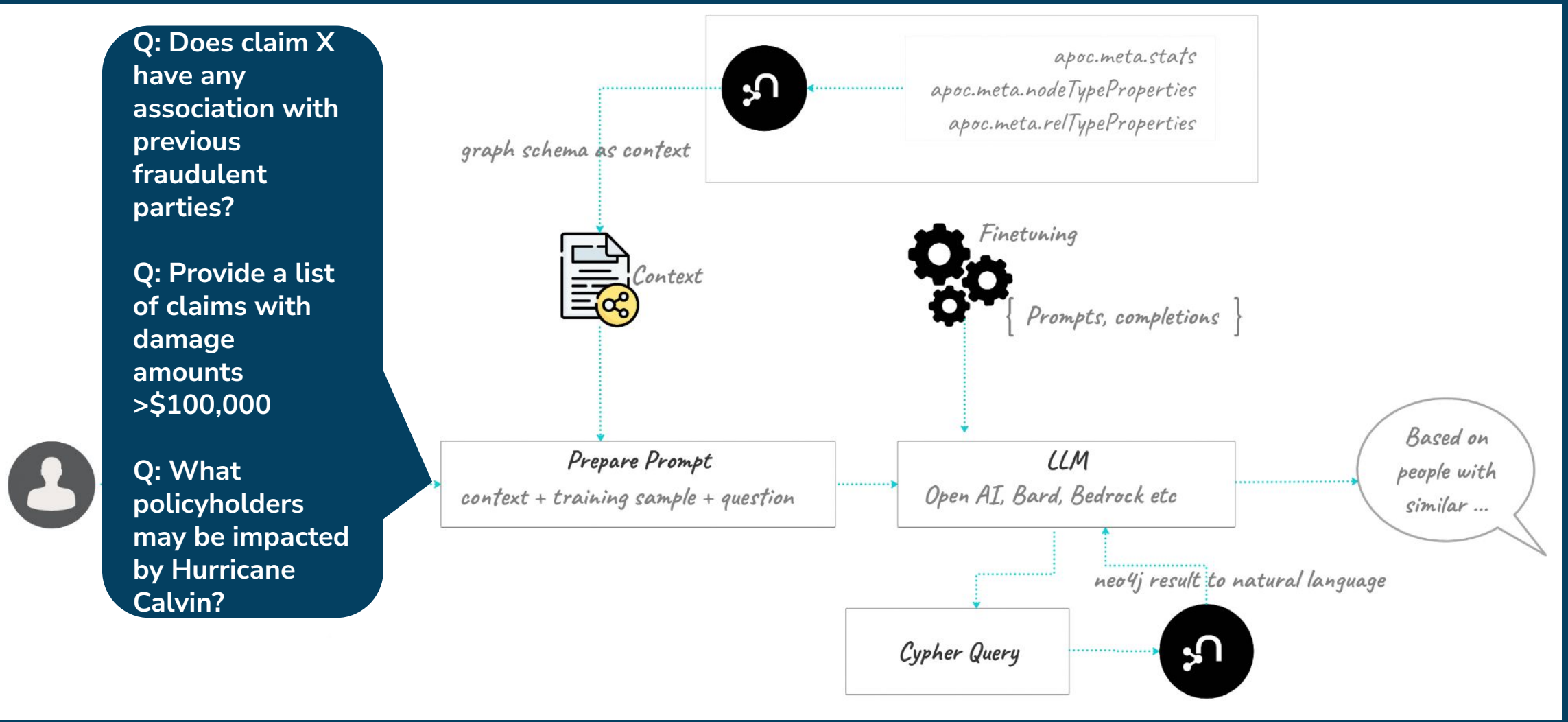
E.g. What is the impact of Hurricane Calvin?

Hurricane Calvin caused minor flooding in Hawaii....

vs...

50 policyholders may be at risk of property damage due to Hurricane Calvin.

Knowledge Retrieval with Neo4j



Demo Intro

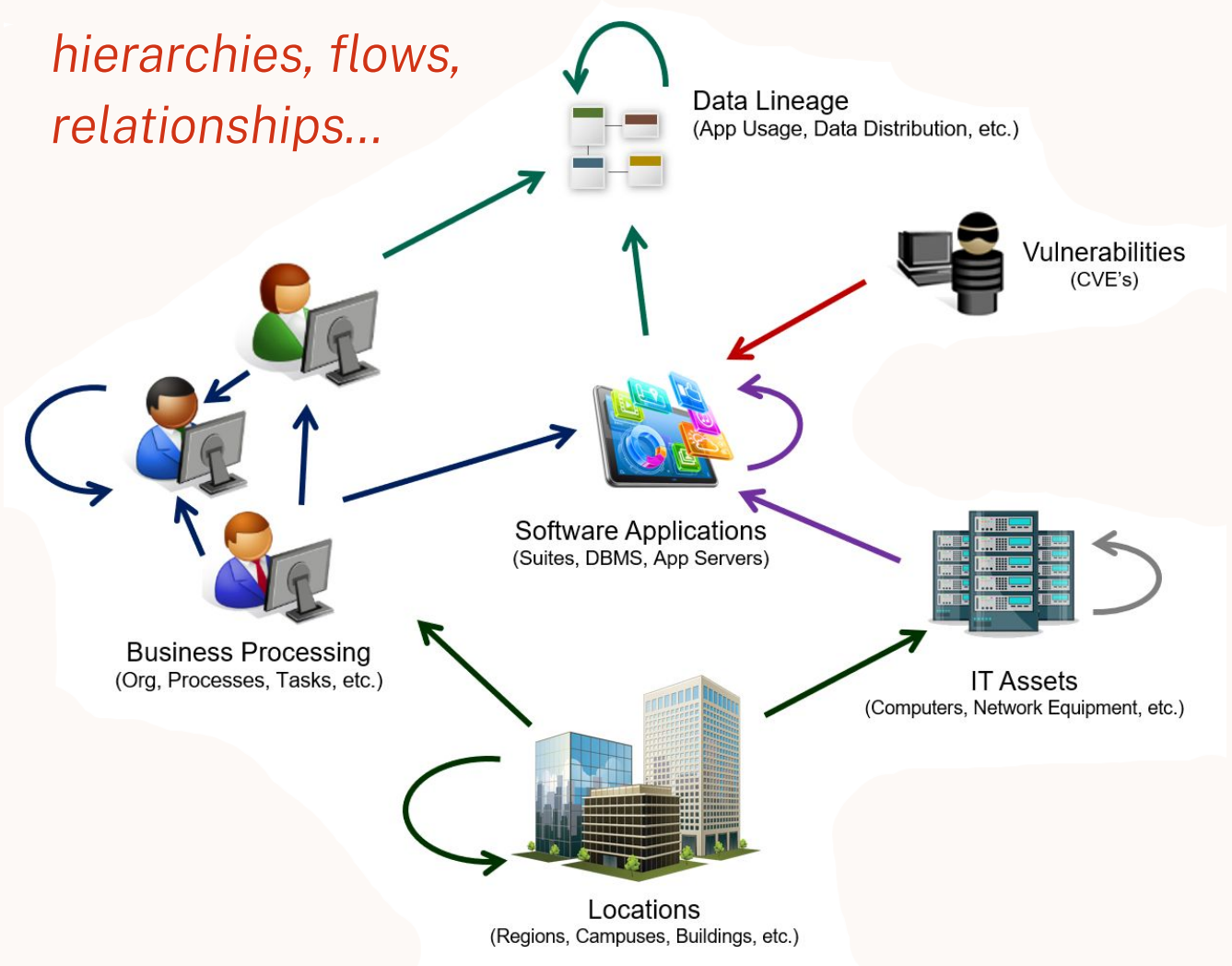
LLM Chatbot

Typical Business Resilience Data

Analyze business impact of

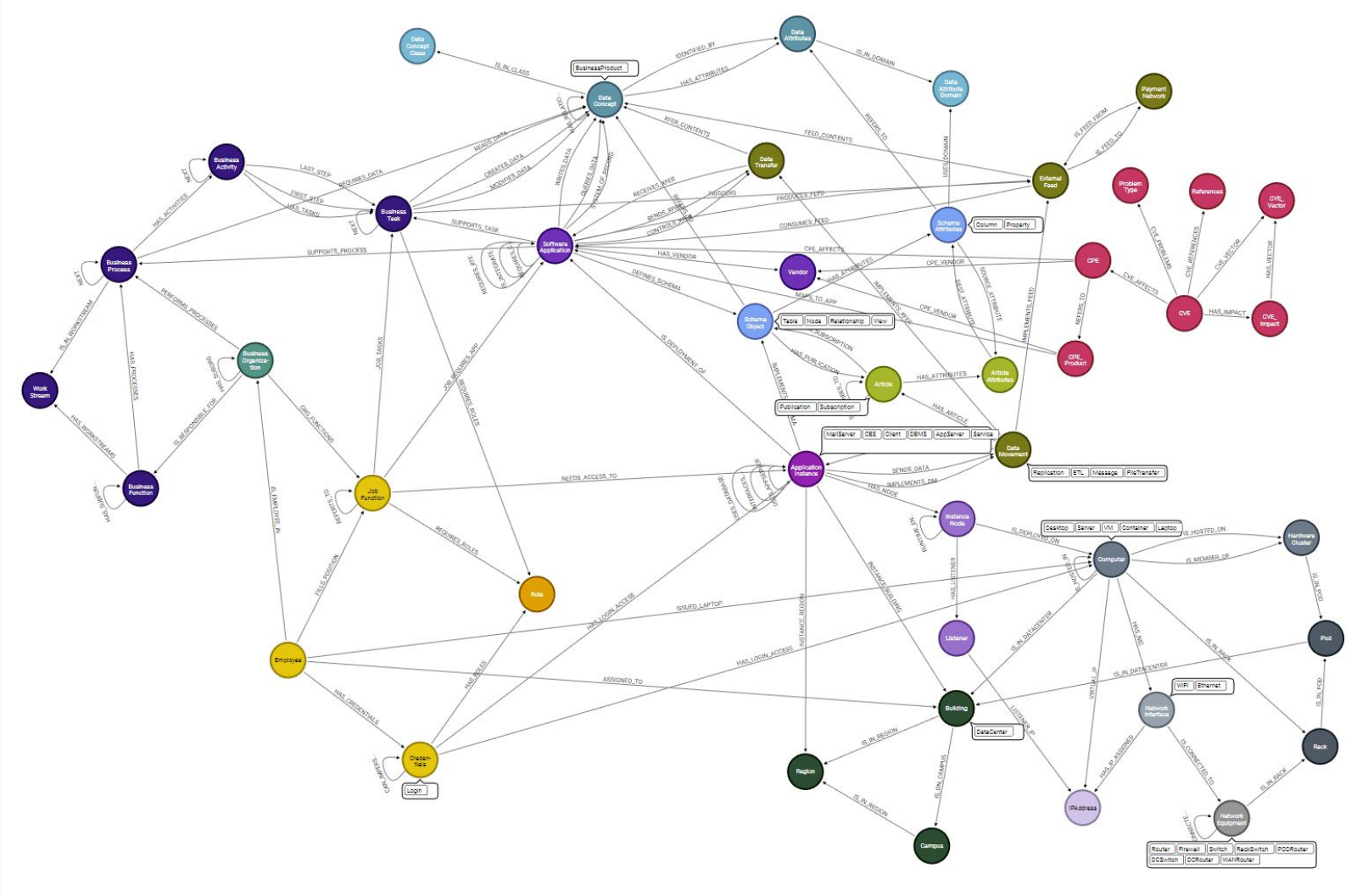
- software & OS vulnerabilities,
- hardware & software upgrades,
- building/geographic disasters
- changes to business data formats

...across mission critical applications and business locations



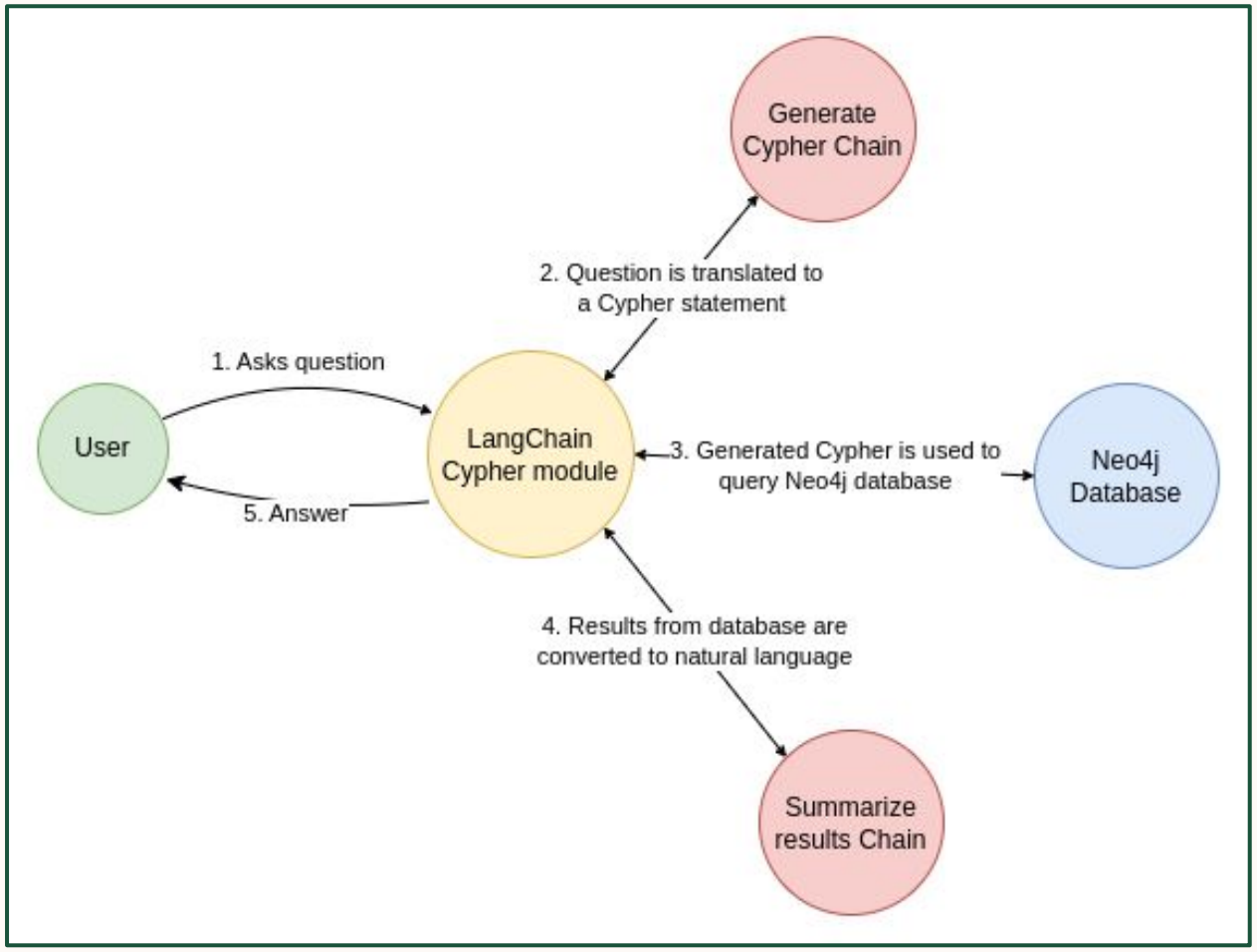
Actual Data Model

CVE Data
Business Data Elements
Vendors,Software
Business Tasks
Application Instances
Data Transfers
People/Roles
Locations
IT Assets

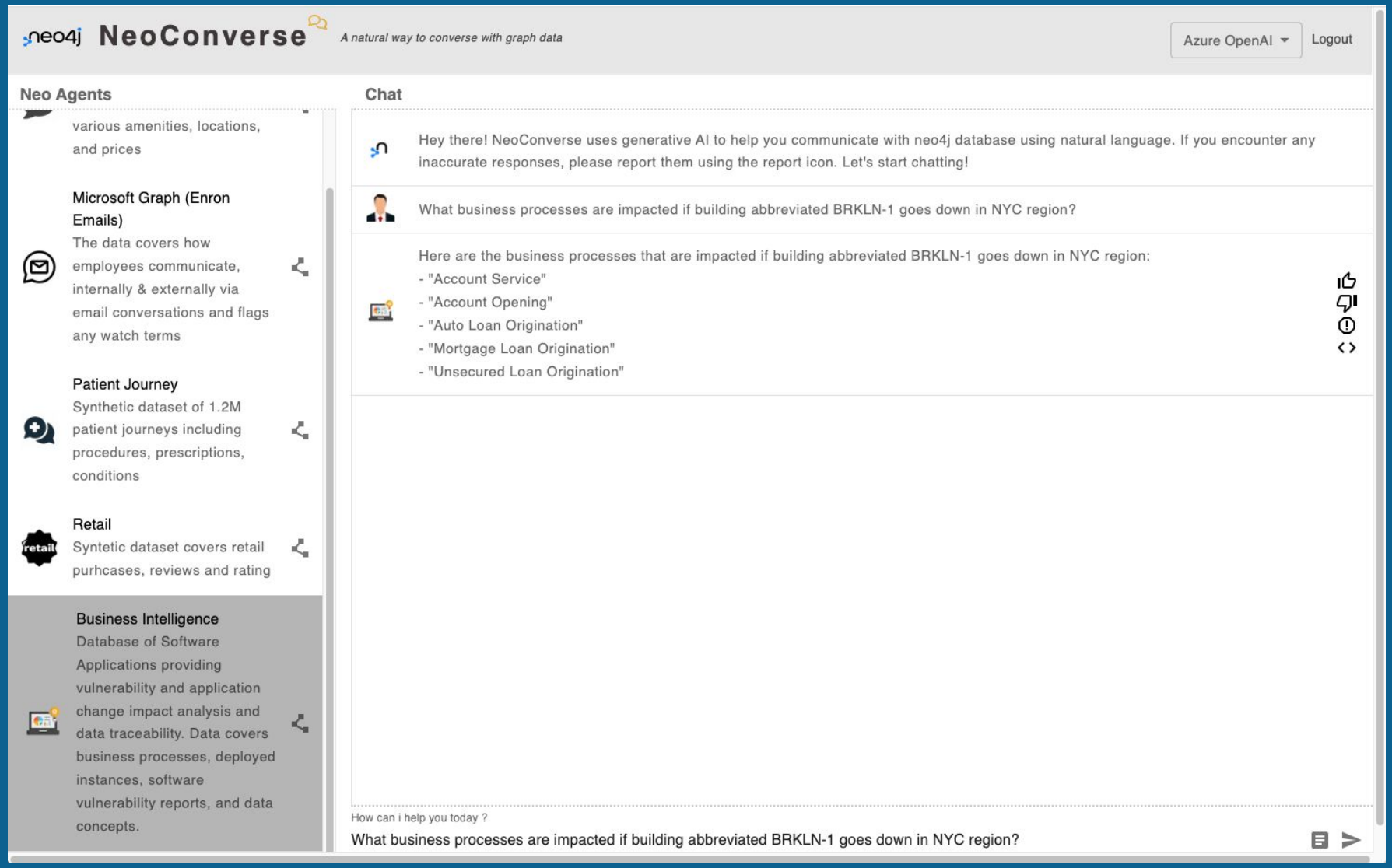


LangChain Demo Application

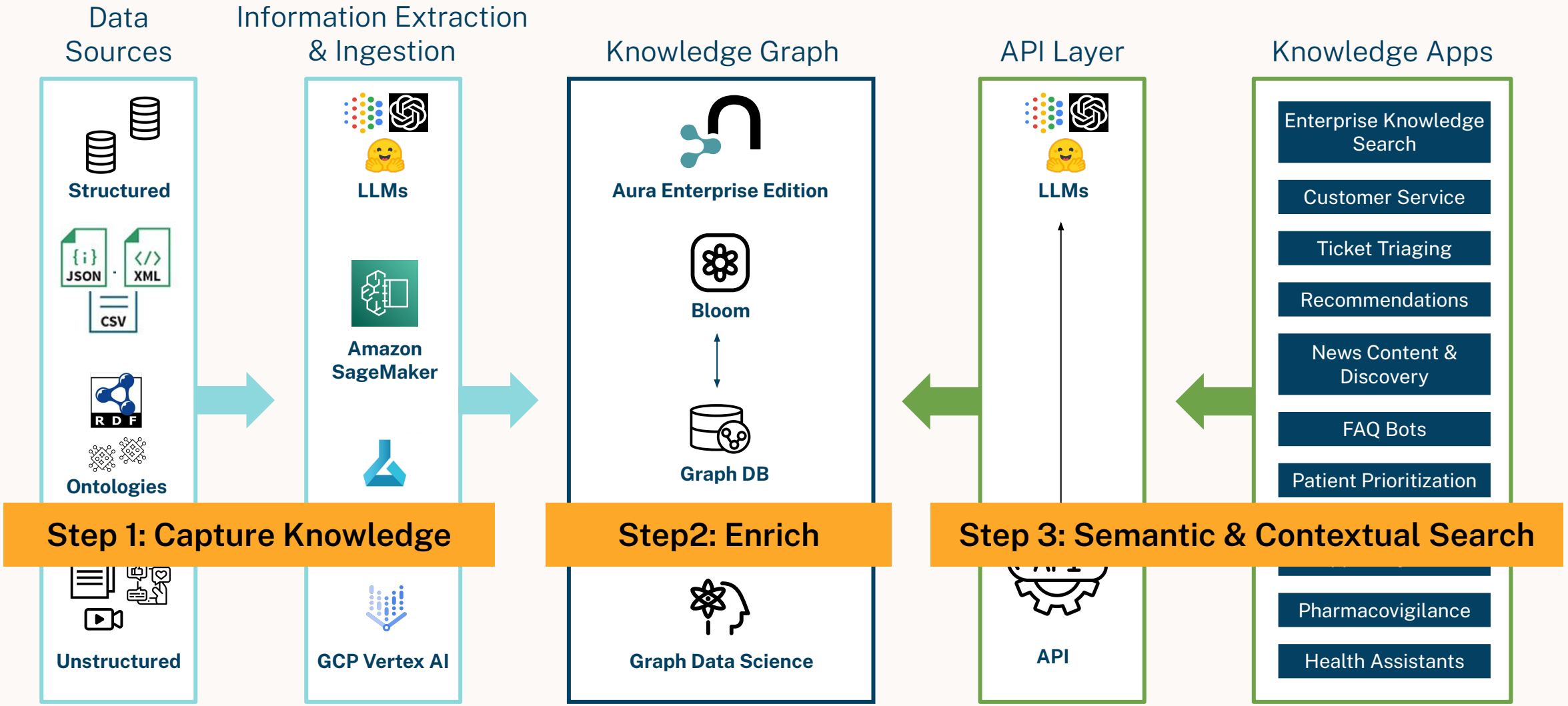
- Translates English to Cypher
- Consumption using LLM model with few shot prompting
- Data augmentation from Neo4j response



Demo



Neo4j Database



Neo4j Enriches All Phases of an AI Ecosystem

Connect

Knowledge Graphs

Analyze

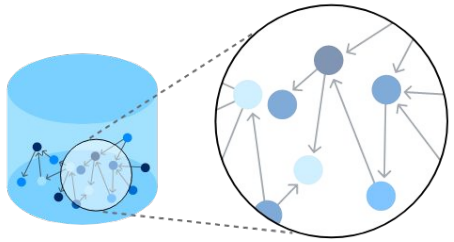
Graph Feature Engineering and Graph ML

Interpret

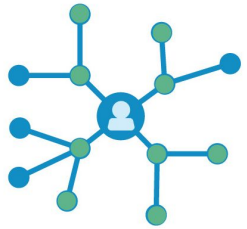
Graph Analytics, Investigations and Counterfactuals

Capitalize

Integrations and Knowledge Graphs for Heuristic AI



Neo4j



Neo4j GDS

Neo4j Bloom

Neo4j Connectors

Thank You!

katie.roberts@neo4j