



Knowledge Graphs & GraphRAG

Essential Ingredients for Enterprise GenAI

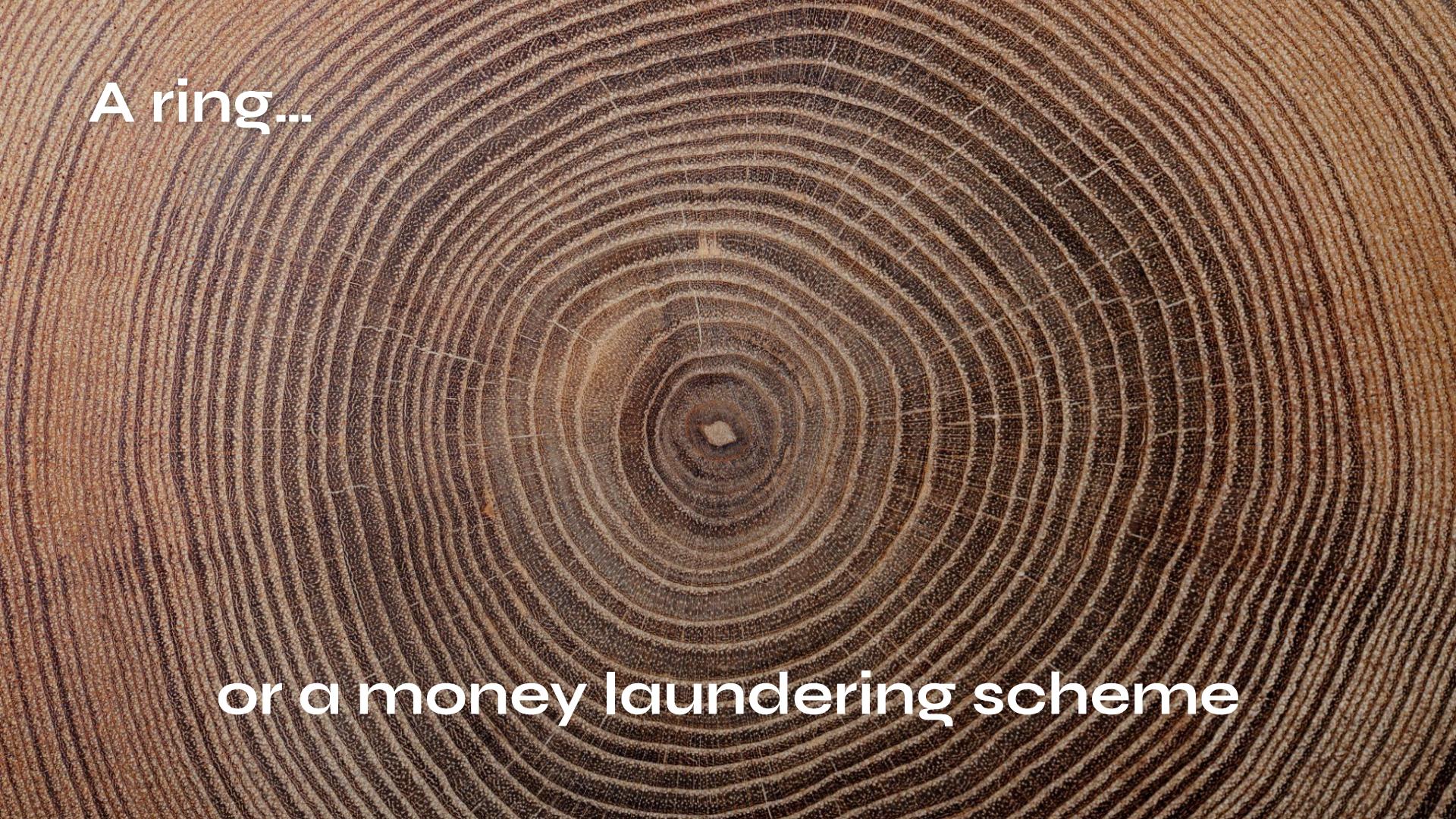
Dr. Jesús Barrasa | Head of Solutions Architecture, Neo4j

TODAY'S REALITY

At the heart of every enterprise challenge is an **explosion of data complexity**



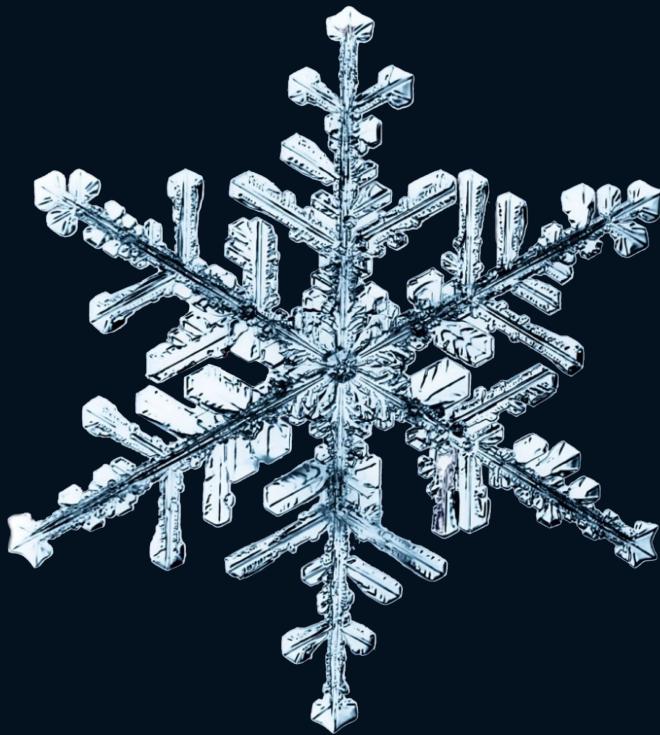
**Explore the connections in your
data to unlock deeper insights**

The background image is a close-up, circular cross-section of a tree trunk. It features numerous concentric growth rings, which are darker in color at the center and become lighter towards the outer edges. The rings are clearly defined by the annual growth increments. A small, light-colored knot hole is visible near the top left of the central area.

A ring...

or a money laundering scheme

A star...



or a critical device in a network



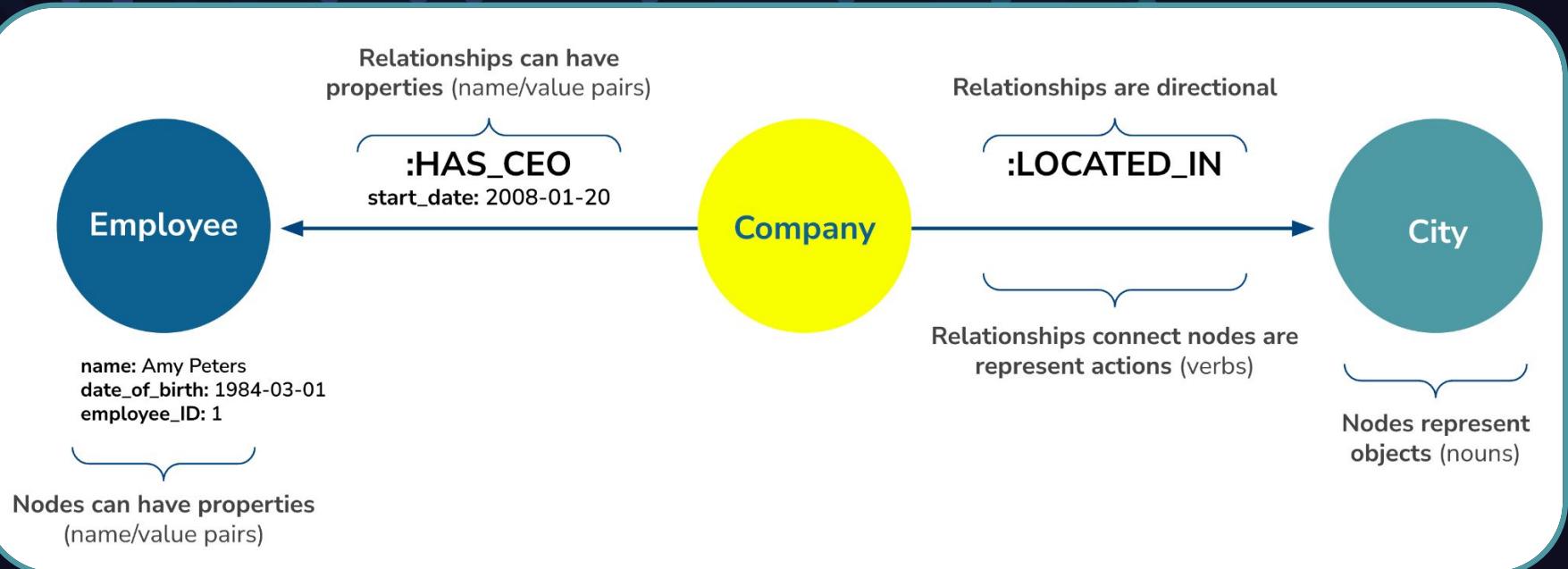
A cloud...

or a community of people



A fundamentally new approach is needed

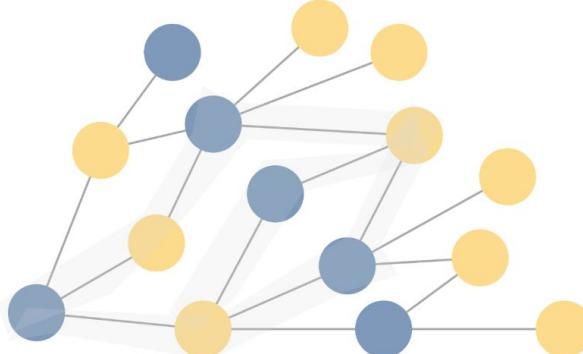
The Property Graph: Simply Powerful



Cypher (GQL): Pattern Based

Node
Pattern

```
MATCH (p:Employee {employee_ID: 1})-[r:WORKS_AT*..3]-(c:Company)  
RETURN c.name as company, count(*) as strength  
ORDER BY strength DESC
```



Relationship
Pattern

ISO GQL '24: New ISO DB Language

“ ISO standardization of Graph Query Language (GQL) ensures that property graph handling will remain open as a de jure standard, similarly to SQL.”

— James Governor, Principal Analyst and Co-Founder, RedMonk

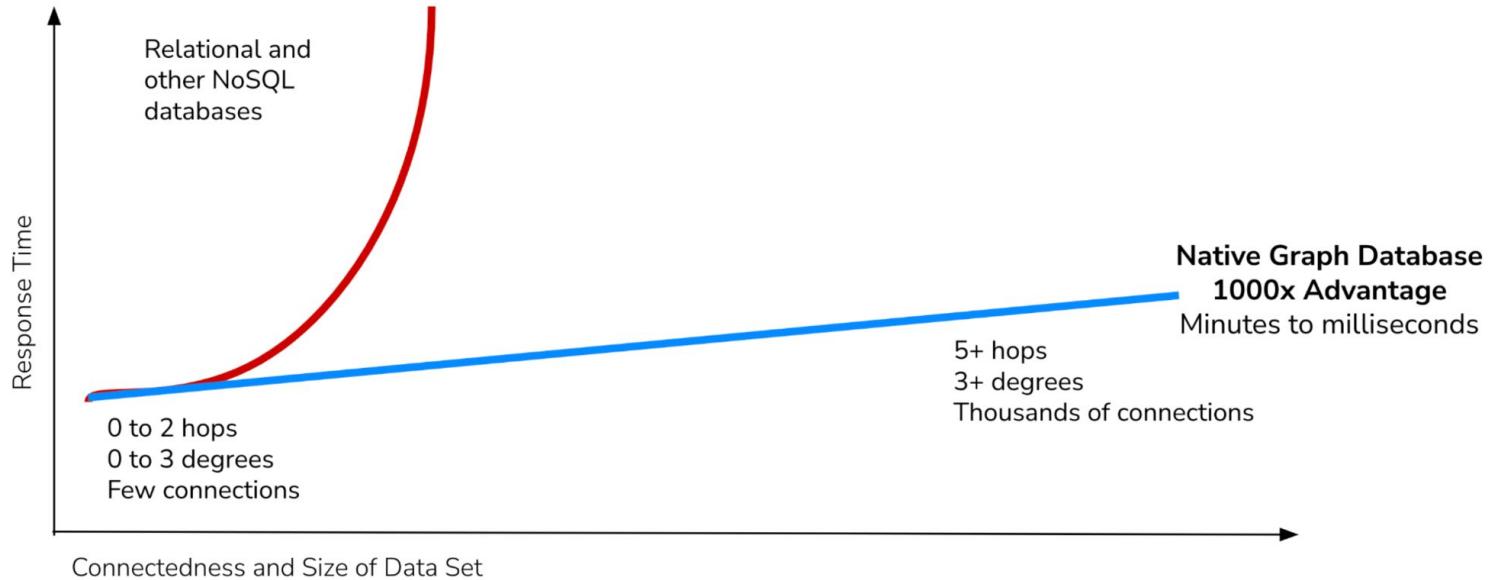
“ It marks a new chapter in the history of query languages, revolutionizing the way we interact with and harness the power of interconnected data.”

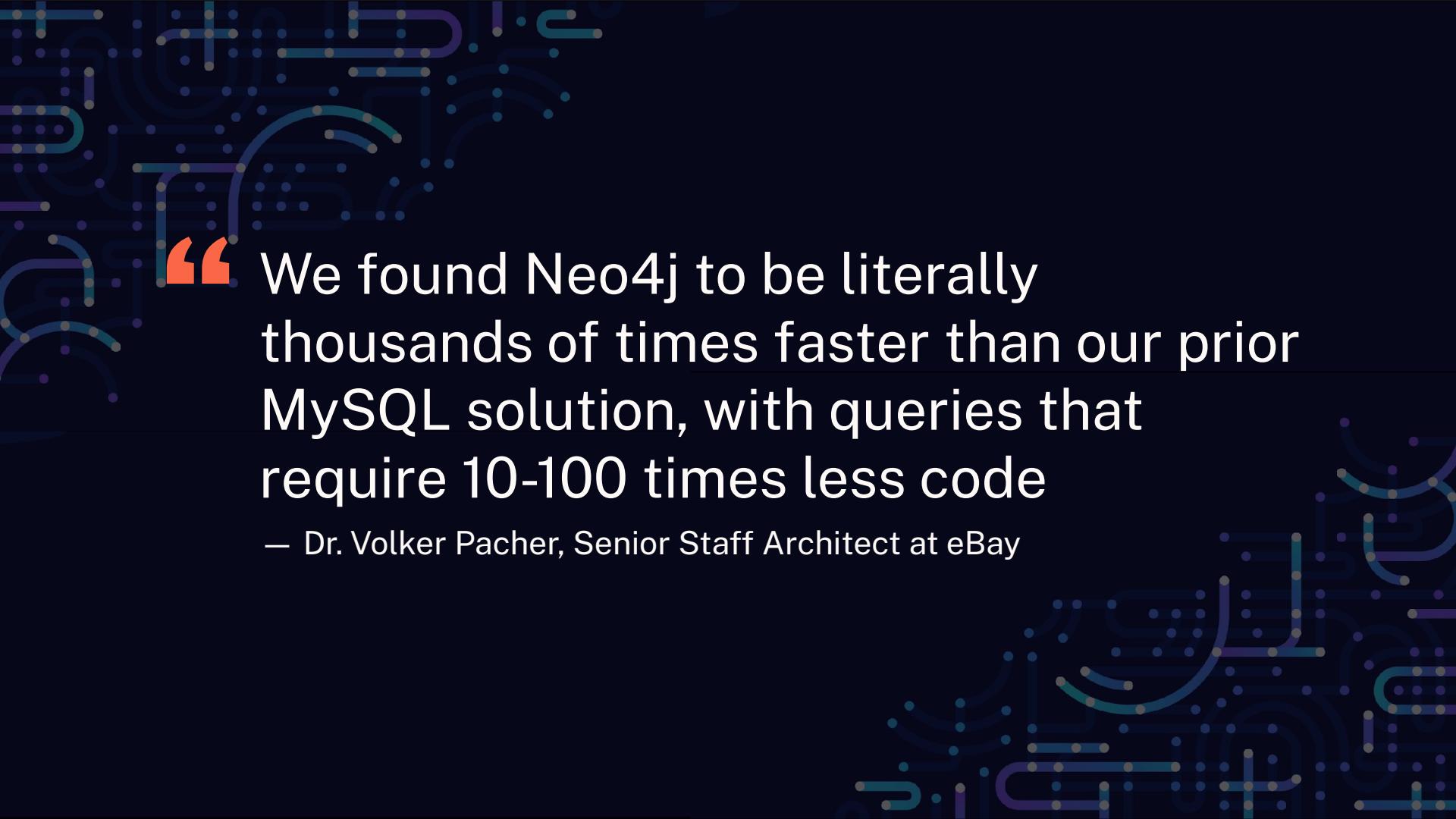
— Tony Holland, Communications Lead, ISO/ IEC Joint Technical Committee 1/AG 1

<https://www.iso.org/standard/76120.html>



Performance: Graph Native Storage





“ We found Neo4j to be literally thousands of times faster than our prior MySQL solution, with queries that require 10-100 times less code

— Dr. Volker Pacher, Senior Staff Architect at eBay



Supply chain resilience



Customer experience



Digital twin



Real-time fraud detection



Drug discovery



can you explain why are they south of Gold Coast?

and Perth?

Perth, located on the western coast of Australia, is not more southern than the Gold Coast. While it is a major city with a population exceeding a million, Perth's latitude is approximately 31.9° S, which actually places it north of the Gold Coast's latitude of about 28.0° S. Therefore, Perth is more northern compared to the Gold Coast, contrary to Melbourne, Adelaide, and Sydney which are all further south.

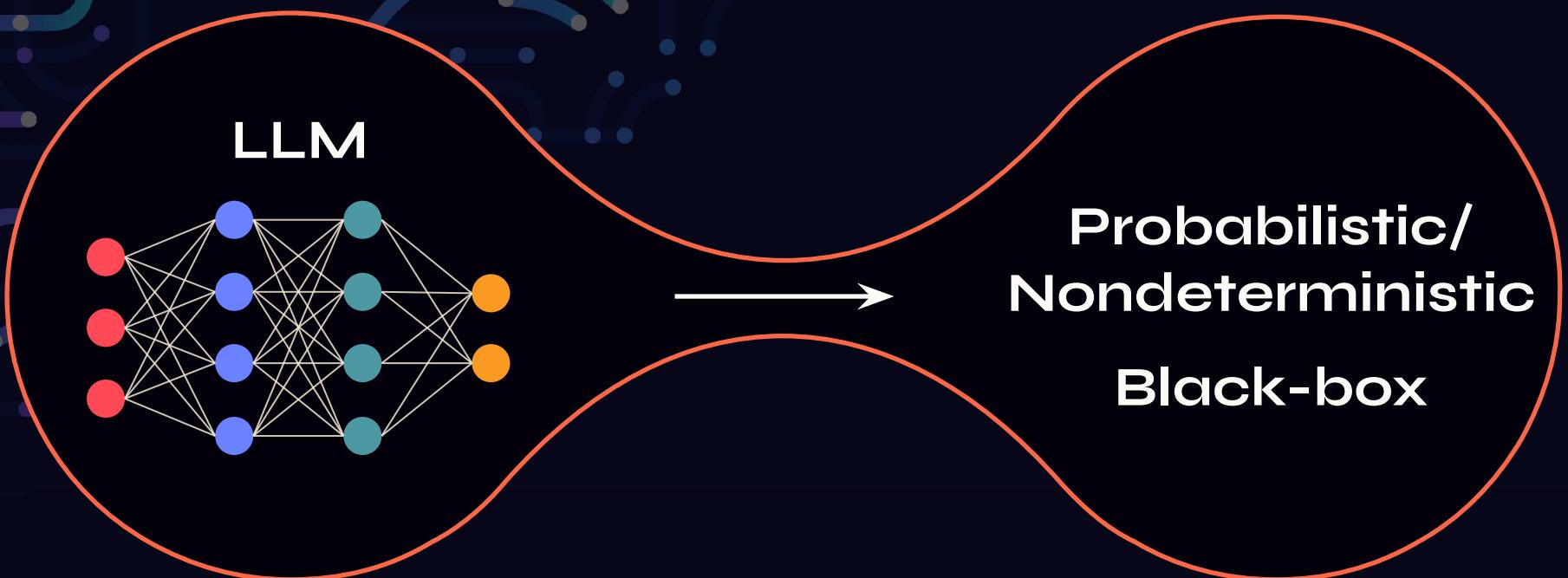
$28.0 > 31.9 ???$

The reason for this counterintuitive result lies in the way latitude is measured. Latitude measures how far north or south a location is relative to the equator, with higher degrees in either direction indicating proximity towards the poles. Hence, both Melbourne and Adelaide, having higher (southern) latitudinal values than Gold Coast, are positioned further south on the Australian continent.

“Houston, we have a problem”

- Hallucinations
- General reasoning vs Domain Reasoning
- Ethical and bias concerns
- Access to the most recent data
- Inability to verify answers or do attribution (!!)
- Lack of enterprise domain knowledge
- Privacy & security challenges

Why are LLMs hitting a barrier?



GPT-3.5 marked the beginning of a
GenAI hockey stick moment



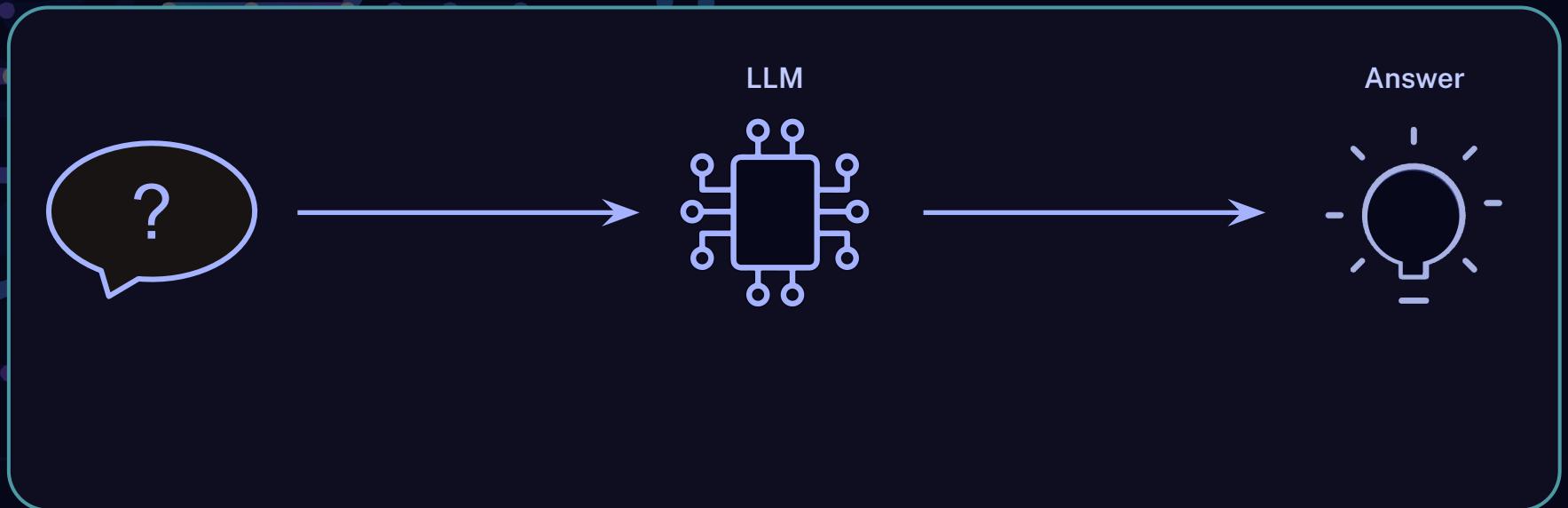
GenAI Go-to-production Barrier

LLMs



OpenAI

So how do we overcome the limitations of LLMs?

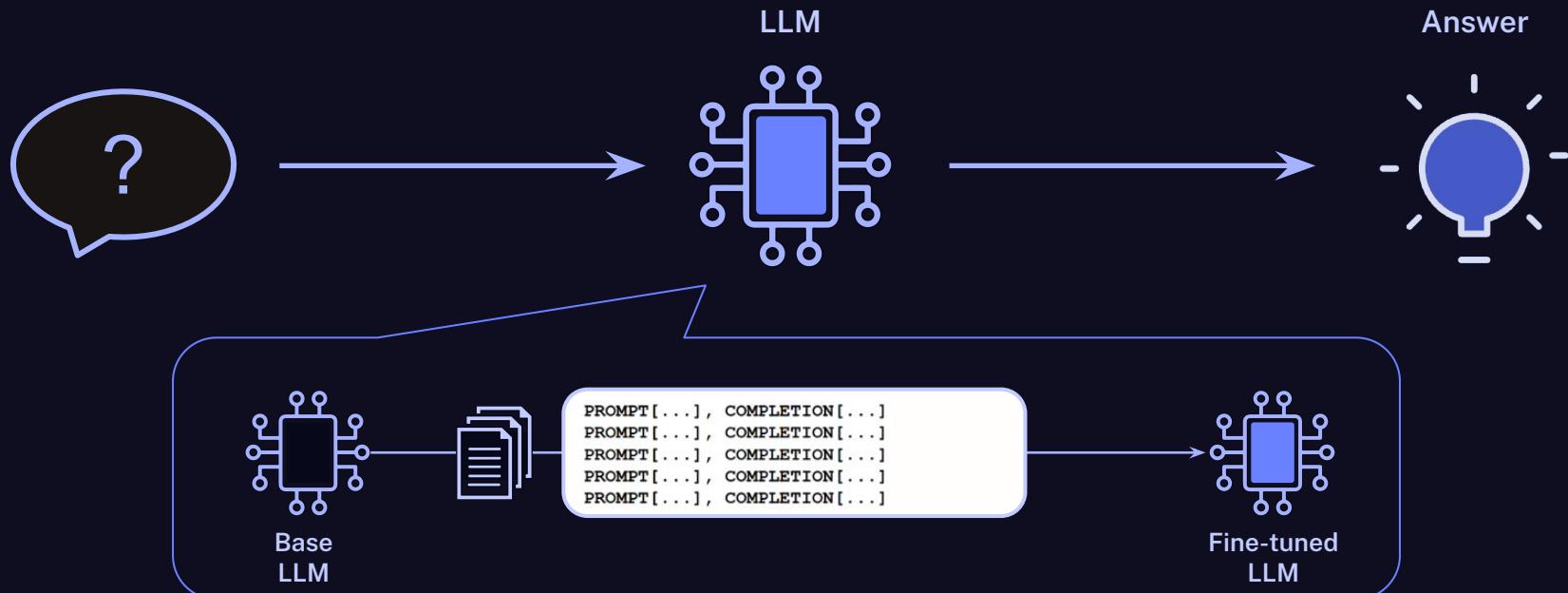




How can we go
further?

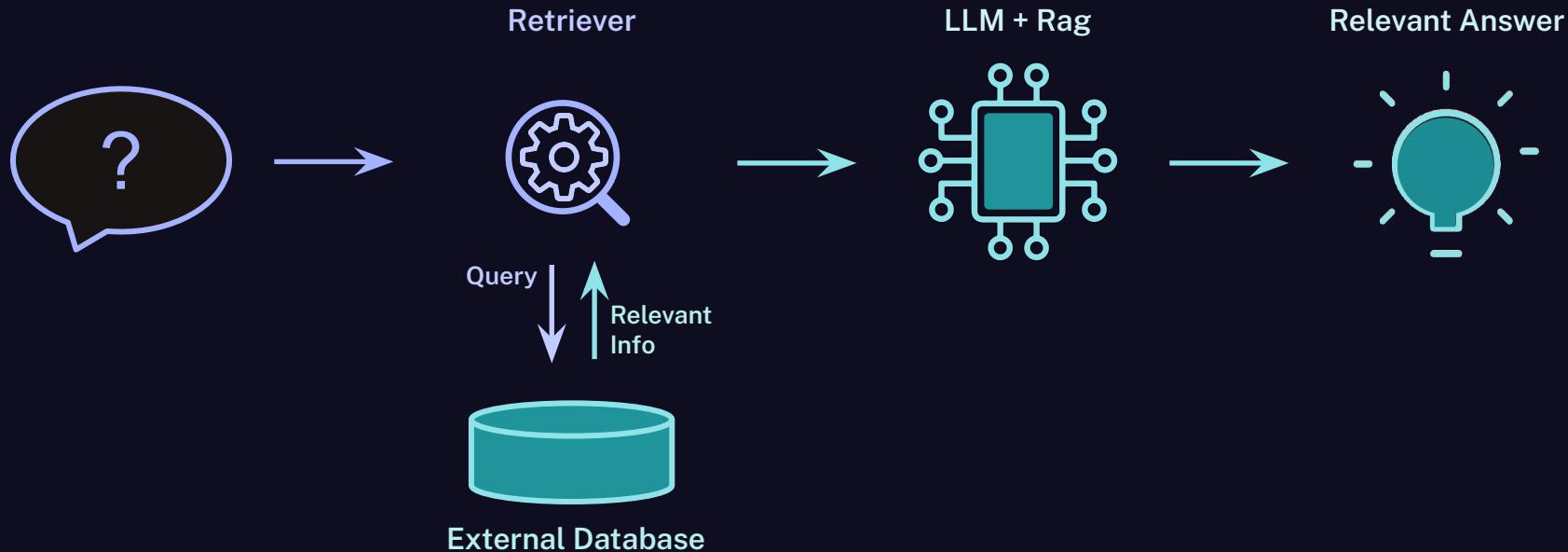
#1 Fine-Tuning

Better task-specific answers from existing knowledge



#2 RAG

Retrieval-Augmented Generation



RAG works

It raises the ceiling, but it's still there

↑
LLMs with
fine-tuning



OpenAI



MISTRAL
AI



ANTHROPIC



RAG



} Go-to-Production
use cases unlocked
with RAG





So what's Next?

Fine
tuning

RAG

?

neo4j

Knowledge Graphs &
GraphRAG
Essential Ingredients

Dr. Jesús Barrasa | Head of Product

GraphRAG

What is GraphRAG?

GraphRAG is **RAG where the R path includes a Knowledge Graph**





Knowledge Graph: What's the Big Deal?

Google The Keyword Latest stories Product updates Company news

SEARCH

Introducing the Knowledge Graph: things, not strings

May 16, 2012 · 4 min read

Amit Singhal
SVP, Engineering

<https://blog.google/products/search/introducing-knowledge-graph-things-not/>

The image shows a Google search results page for the query "sagrada familia". The results include a map, several images of the Sagrada Família, and a detailed description of the basilica.

Map: A map showing the location of La Sagrada Família in Barcelona, Spain, with a red marker indicating the exact location.

Image: A large image of the Sagrada Família's exterior.

Text: "La Sagrada Família" Basilica in Barcelona, Spain

Rating: 4.7 stars from 214,752 reviews

Buttons: Website, Directions, Save, Call, SEE TICKETS

Description: The Basilica i Temple Expiatori de la Sagrada Família, shortened as the Sagrada Família, is an unfinished church in the Eixample district of Barcelona, Catalonia, Spain. It is the largest unfinished Catholic church in the world. [Wikipedia](#)

Address: C/ de Mallorca, 401, 08013 Barcelona, Spain

Hours: Opens soon · 9 am · Updated by this business 3 weeks ago

Phone: +34 932 08 04 14

Construction started: 19 March 1882

Opened: 7 November 2010

Architects: Antoni Gaudí, Jordi Bonet i Armengol, MORE

Height: 172 m

Architectural styles: Modernisme, Art Nouveau Architecture, Gothic Revival architecture, Noucentisme, Spanish Gothic architecture

Province: Barcelona

Function: Church



La Sagrada Familia

Website Directions Save Call

4.7 ★★★★★ 214,752 Google reviews ⓘ

Basilica in Barcelona, Spain

SEE TICKETS

The Basílica i Temple Expiatori de la Sagrada Família, shortened as the Sagrada Família, is an unfinished church in the Eixample district of Barcelona, Catalonia, Spain. It is the largest unfinished Catholic church in the world. [Wikipedia](#)

Address: C/ de Mallorca, 401, 08013 Barcelona, Spain

Hours: Opens soon · 9 am ⓘ

Updated by this business 3 weeks ago

Phone: +34 932 08 04 14

Construction started: 19 March 1882

Opened: 7 November 2010

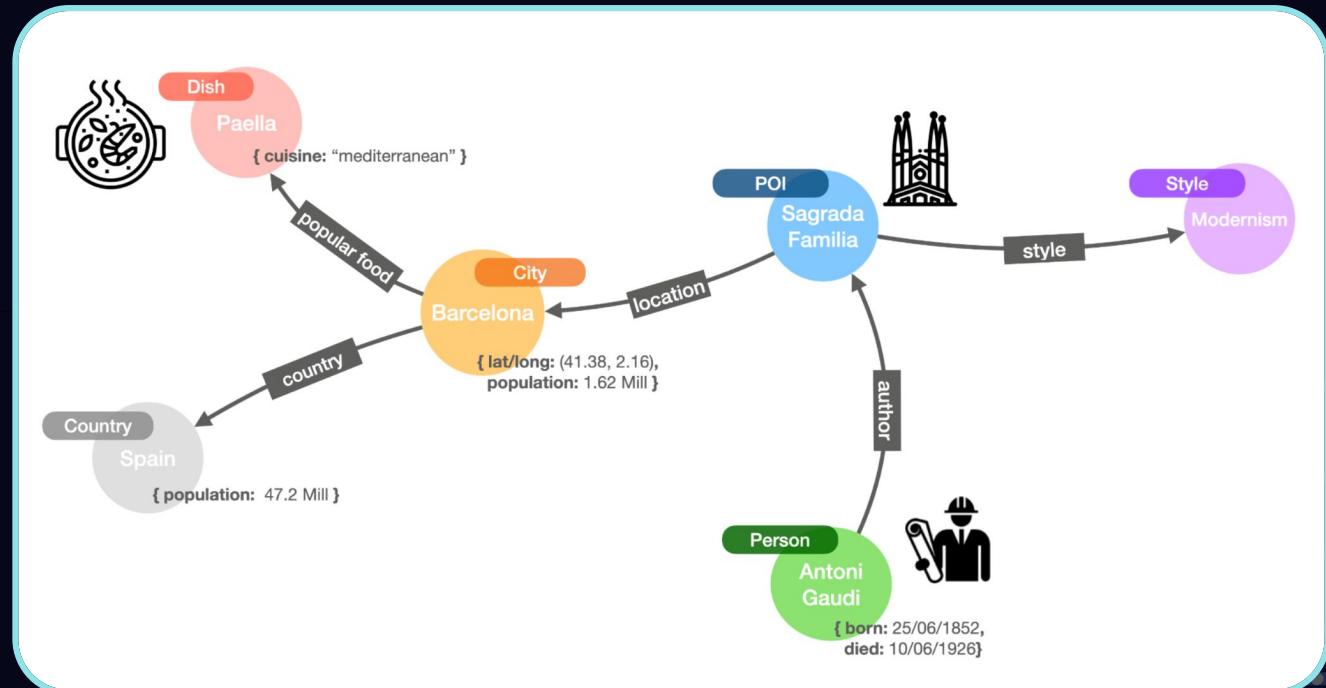
Architects: Antoni Gaudí, Jordi Bonet i Armengol, MORE

Height: 172 m

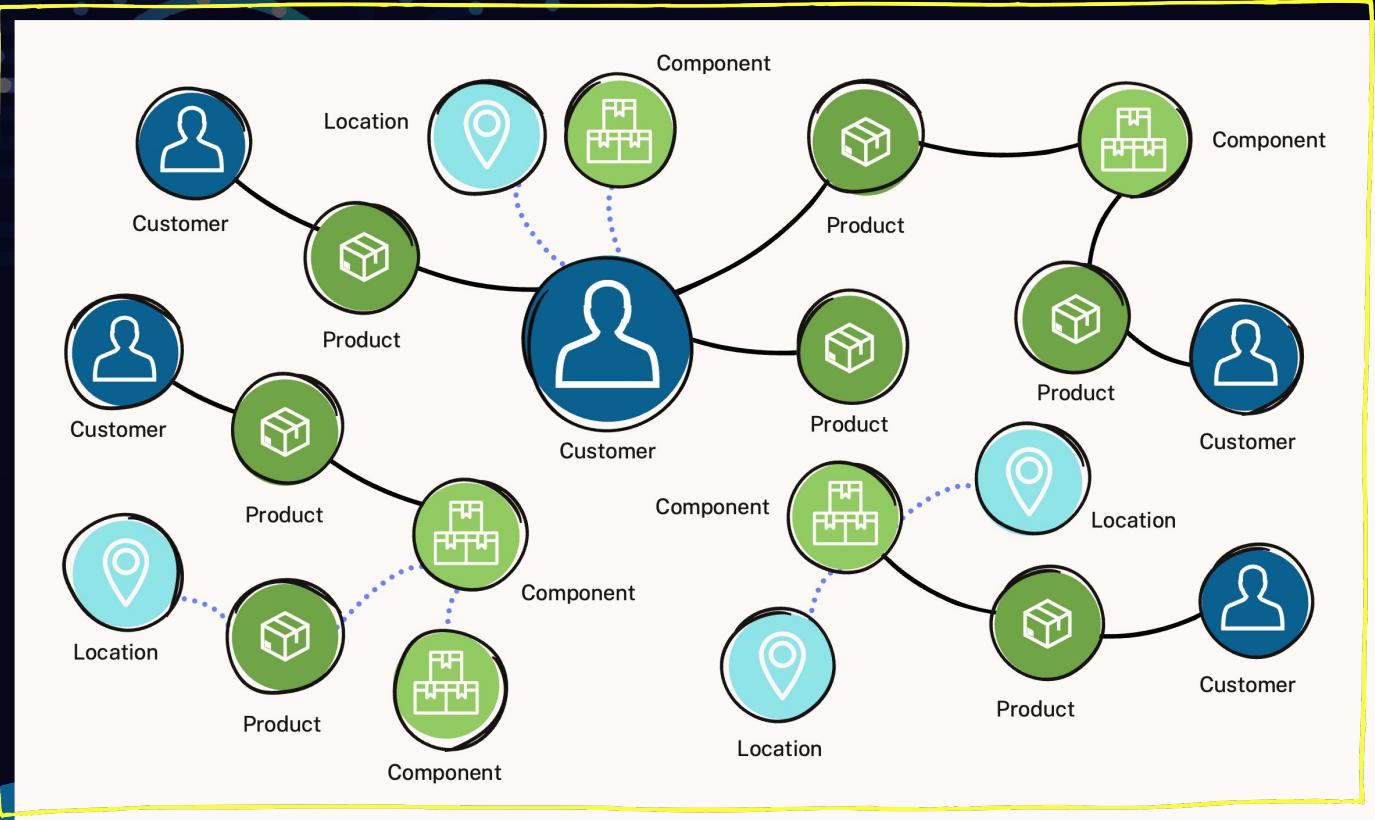
Architectural styles: Modernisme, Art Nouveau Architecture, Gothic Revival architecture, Noucentisme, Spanish Gothic architecture

Province: Barcelona

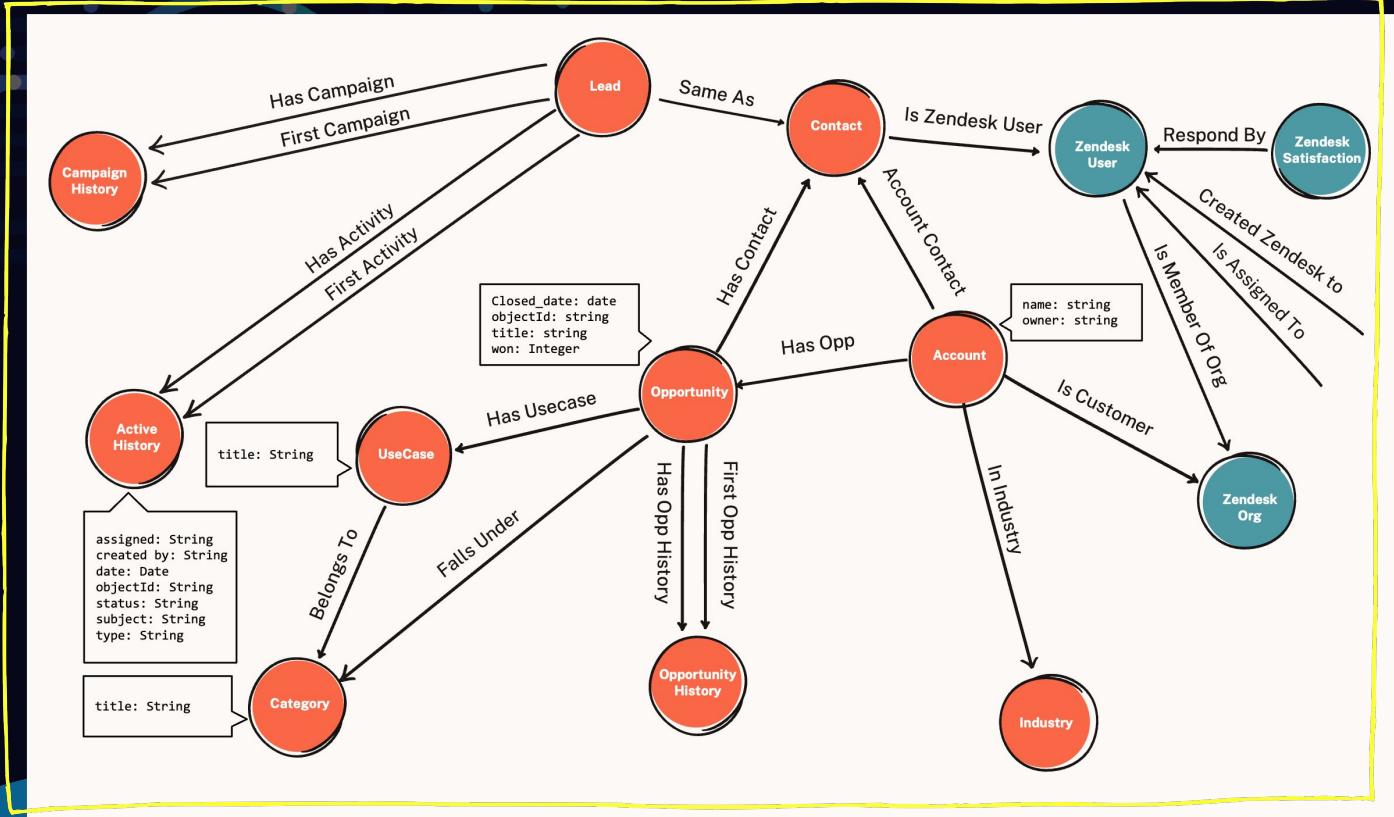
Function: Church



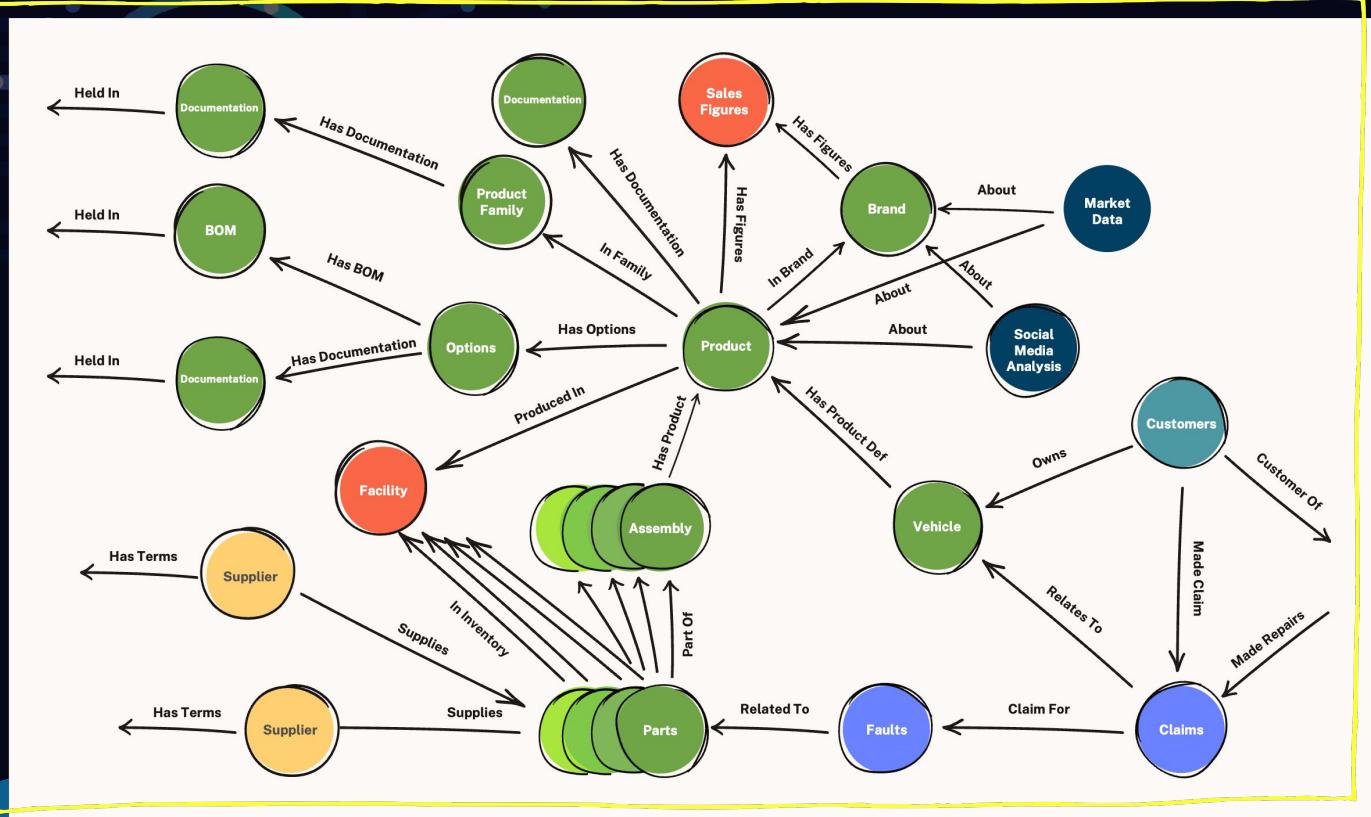
Commerce Transactions Graph



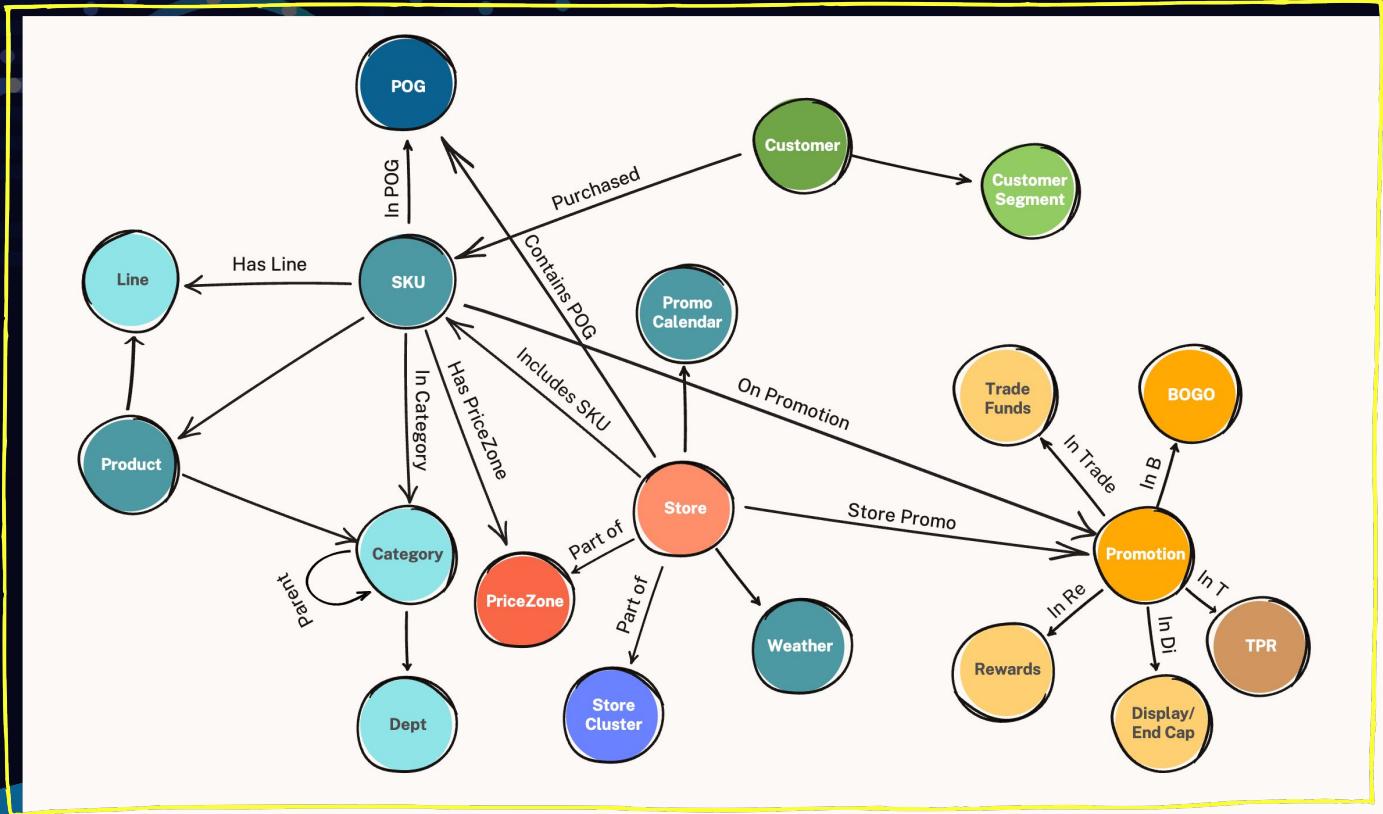
Customer Relationship Graph

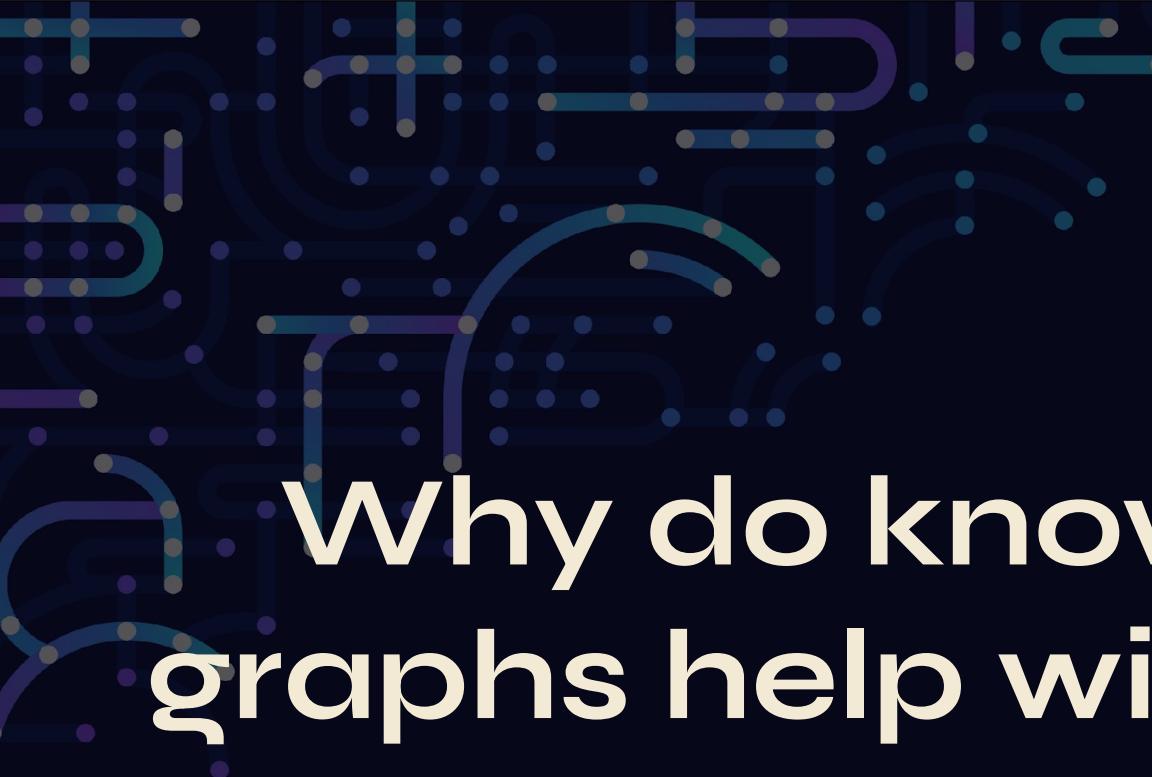


Digital Twin Graph



Product and Parts Graph





Why do knowledge graphs help with GenAI

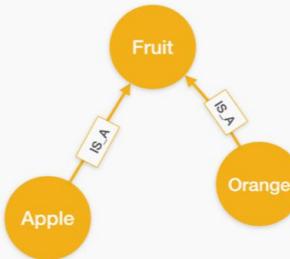
DATA SEMANTICS

EXPLICIT (SYMBOLIC)

IMPLICIT (SUB-SYMBOLIC)



Representation



Representation

Apple	[0.2435, 3.7652, 0.00234, 456.66, ...]
Orange	[115.124, 29.7652, 4.2131, 2.431, ...]
Fruit	[0.0035, 17.661, 0.0113, 11.4566, ...]

DATA SEMANTICS

EXPLICIT (SYMBOLIC)

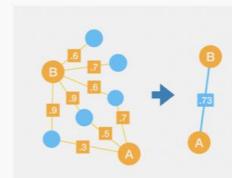
IMPLICIT (SUB-SYMBOLIC)



Similarity calculation

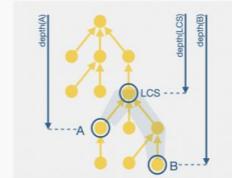
Structural

- ▶ Node similarity
- ▶ Overlap
- ▶ Jaccard



Taxonomy based

- ▶ Path
- ▶ Leacock-Chodorow
- ▶ Wu-Palmer

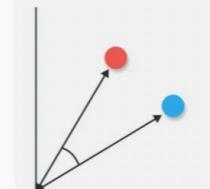


Similarity calculation

Euclidean



Cosine



Vector distance based

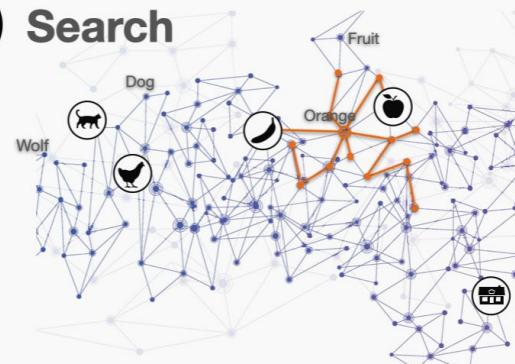
DATA SEMANTICS

EXPLICIT (SYMBOLIC)

IMPLICIT (SUB-SYMBOLIC)

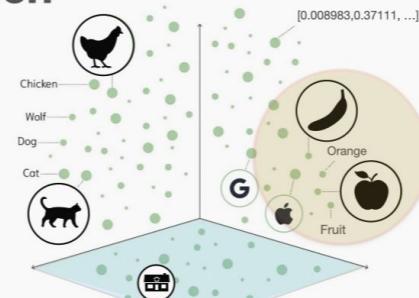


Search



Search

Indexing:
► IVF
► HNSW



GraphRAG unlocks AI

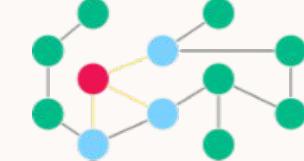
LLMs



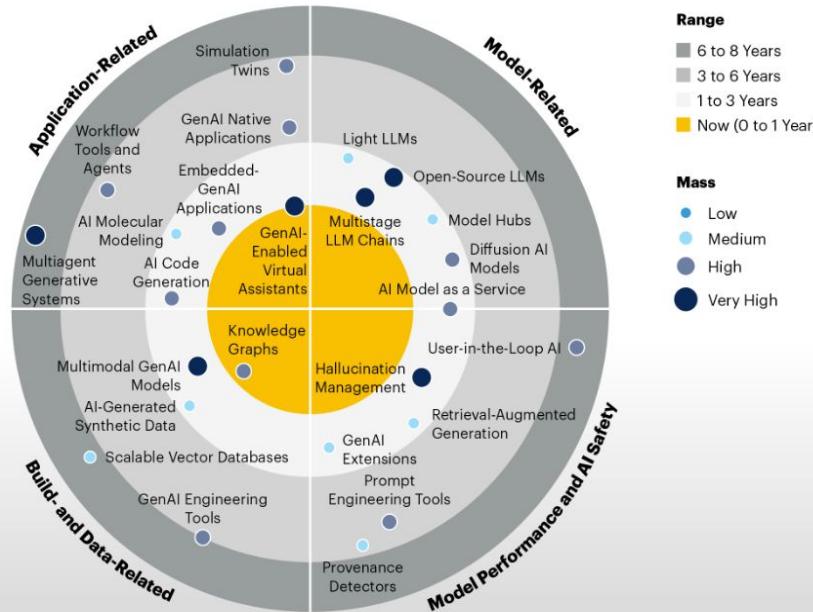
RAG with Vector



GraphRAG



Impact Radar for Generative AI



Source: Gartner
© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. 2683355

Gartner®



<https://www.gartner.com/en/articles/understand-and-exploit-gen-ai-with-gartner-s-new-impact-radar>

A fantastic success story

“...we've started **using extensively in house is Neo4j and graphs**. And we've also looked at how people collaborate on building great information

“And then on top of that we have Kiki, who then **explores that information and brings it to life**. So we can go and ask Kiki about anything!

“And all of that is getting **centralized into one place and connected through the knowledge graph**. We're seeing that that is having a **tremendous impact on productivity** internally. So Kiki is basically our own internal chatbot based on that growing internal knowledge graph.



E 6 TRAINING DATA

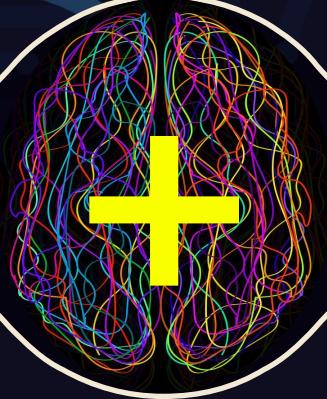
S E B A S T I A N	S I E M I A T K O W S K I
K L A R N A	

Co-founder and CEO Sebastian Siemiatkowski tells how Klarna shipped its customer service assistant and how experimenting with AI is transforming the company.

Read Transcript

● ● ● ● ●

A small rectangular portrait of Sebastian Siemiatkowski, wearing headphones and looking towards the camera, is positioned in the bottom right corner of the slide's footer area.



KG

LLM

Better together



The Benefits of GraphRAG



#1: Answer Quality

Answers are more accurate, but
also more useful & complete



data.world

3x
improvement

A Knowledge Graph improved the accuracy of LLM responses by 54.2%, an average of 3x.



Juan Sequeda · 1st

Principal Scientist & Head of AI Lab at data.world; co-host of Cata...
3mo · ②

Investing in Knowledge Graph provides higher accuracy for LLM-powered question-answering systems. And ultimately, to succeed in this AI world, enterprises must treat the business context and semantics as a fit ...see more

A BENCHMARK TO UNDERSTAND THE ROLE OF KNOWLEDGE GRAPHS ON LARGE LANGUAGE MODEL'S ACCURACY FOR QUESTION ANSWERING ON ENTERPRISE SQL DATABASES

TECHNICAL REPORT

Juan F. Sequeda
data.world
juan@data.world

Dean Allemann
data.world
dean.allemann@data.world

Bryon Jacob
data.world
bryon@data.world

November 14, 2023

ABSTRACT

Enterprise applications of Large Language Models (LLMs) hold promise for question answering on enterprise SQL databases. However, the context in which LLMs can accurately respond to enterprise questions in such databases remains unclear, given the absence of suitable Text-to-SQL benchmarks tailored to enterprise settings. Additionally, the potential of Knowledge Graphs (KGs) to enhance LLM-based question answering by providing business context is not well understood. This study aims to evaluate the accuracy of LLM-powered question answering systems in the context of enterprise SQL databases, while also exploring the role of knowledge graphs in improving accuracy. To achieve this, we created a benchmark dataset involving an enterprise SQL database, the AutoInsure in the insurance domain, a range of enterprise queries encompassing reporting to metrics, and a contextual layer incorporating an ontology and mappings that define a knowledge graph. Our primary finding reveals that question answering using GPT-4 with zero-shot prompts directly on SQL databases, achieves an accuracy of 16%. Notably, this accuracy increases to 54% when questions are posed via a Knowledge Graph, and to 80% when the Knowledge Graph is integrated into the LLM's reasoning process.



dbt Labs

70K followers
3mo · ②

Juan Sequeda and the data.world team dropped a bombshell paper that validates the intuition held by many of us—layering structured Semantic Knowledge on top of your data leads to much stronger ability to correctly answer ad-hoc questions about your organizational data with Large Language Models

So what does it all mean, and how does the dbt Semantic Layer measure up? ↗

Check out Jason Ganz' blog post to find out ↗

The Analytics Engineering



"Across all metrics, our method demonstrates consistent improvements. Notably, it surpasses the baseline by 77.6% in MRR and by 0.32 in BLEU score, substantiating its superior retrieval efficacy and question-answering accuracy."

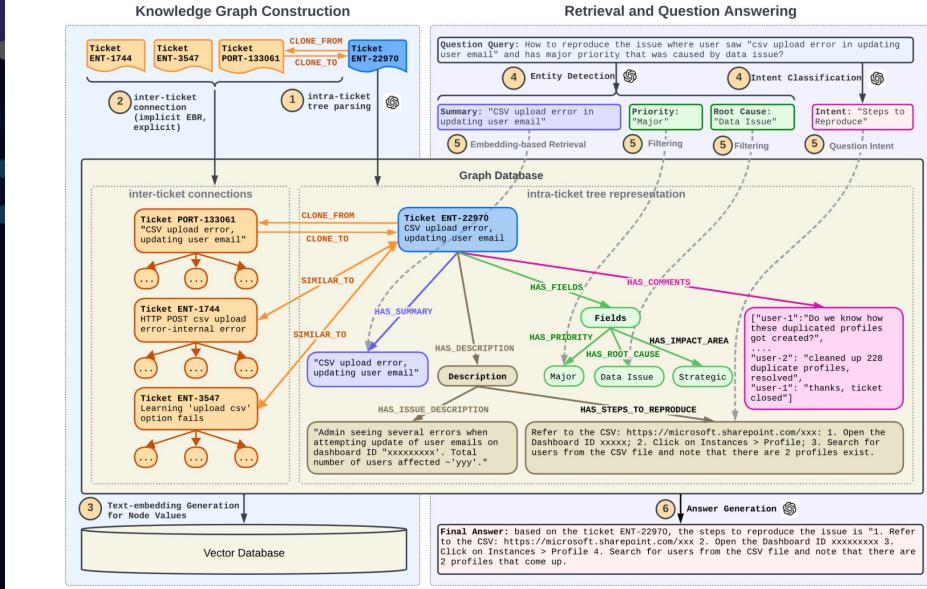


Table 1: Retrieval Performance

	MRR	Recall@K		NDCG@K	
		K=1	K=3	K=1	K=3
Baseline	0.522	0.400	0.640	0.400	0.520
Experiment	0.927	0.860	1.000	0.860	0.946

Table 2: Question Answering Performance

	BLEU	METEOR	ROUGE
Baseline	0.057	0.279	0.183
Experiment	0.377	0.613	0.546

Retrieval-Augmented Generation with Knowledge Graphs for Customer Service Question Answering

Zhenhai Yu shenzheli@tencent.com Tencent Corporation Shenzhen, China	Mark Jermyn Carr markcarr@tencent.com Tencent Corporation Shenzhen, China	Matthew Gavves matthewgavves@tencent.com Tencent Corporation Shenzhen, China
Tie Wang tiwang@tencent.com Tencent Corporation Shenzhen, China	Manasi Deshpande mdeshpande@tencent.com Tencent Corporation Shenzhen, China	Xiaofeng Wang xiaofw@tencent.com Tencent Corporation Shenzhen, China

ABSTRACT
Customer service support, swiftly and accurately retrieving relevant past issues is critical for efficient resolving customer questions (FAQs) for large language models (LLMs). We present a large corpus of past issue tracking tickets in port, and propose a retrieval-augmented generation (RAG) framework for generating responses (FAQs) to large language models (LLMs) for customer service questions. Our method constructs a KG from historical ticket data and uses a knowledge graph (KG) to store the ticket information. This enables the system to quickly retrieve past issues to generate answers. This integration of a KG not only improves the system's ability to handle complex questions, but also enhances its overall quality by integrating the ticket information. Our experiments show that our proposed RAG model achieves state-of-the-art performance on the SQuAD 2.0 dataset, utilizing key metrics (F1, Recall@1, NDCG@1, and MRR@1). The proposed RAG model also outperforms the baseline methods by 77.6% in MRR and by 0.32 in BLEU score. The proposed RAG model can be deployed as a real-time service for generating answers on mobile and has reduced the median per-issue execution time by 24x.

KEYWORDS
Large Language Model, Knowledge Graph, Question Answering, Retrieval-Augmented Generation

ACM Reference Format:
Zhenhai Yu, Mark Jermyn Carr, Matthew Gavves, Tie Wang, Manasi Deshpande, and Xiaofeng Wang. 2024. Retrieval-Augmented Generation with Knowledge Graphs for Customer Service Question Answering. In Proceedings of the 17th International ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24), June 10–12, 2024, Seattle, WA, USA. © 2024, Author(s). Licensee ACM, Inc., New York, NY, USA. 9 pages. <https://doi.org/10.1145/3587287.3587319>

1 INTRODUCTION
Effective individual support in customer service requires problem-solving and question-answering capabilities. A major challenge is how to quickly and accurately retrieve relevant past issues to generate responses. This process typically consists of two stages: first, historical ticket data is used to build a knowledge graph (KG), which stores the context length constraints of embedding models, such as BERT and RoBERTa, and the second stage is to query the KG during the question answering phase, the system retrieves the most relevant answers from the KG and generates the final response. Previous work has shown that using KGs can significantly improve the quality of generated responses [1].

[1] Specifically enhanced retrieval performance and question answering quality by using KGs.

† Contributors: 1. Comprehended Retrieval-free Querying Structures: Issue tracking documents such as JIRA, GitHub, and Trello are widely used in customer service. These systems provide a structured way to store and retrieve information related to specific issues. Despite its straightforward approach, this method often fails to handle complex queries involving multiple entities and their relationships.

‡ Limitations: 1. Comprehended Retrieval-free Querying Structures: Issue tracking documents such as JIRA, GitHub, and Trello are widely used in customer service. These systems provide a structured way to store and retrieve information related to specific issues. Despite its straightforward approach, this method often fails to handle complex queries involving multiple entities and their relationships.

2. Limited Coverage: The proposed RAG model is limited to handling customer service questions related to specific domains, such as IT support and customer support. It may not be able to handle questions from other domains, such as healthcare or finance, without significant domain-specific knowledge.

3. Resource Intensive: The proposed RAG model requires a large amount of historical ticket data to build the KG, which can be computationally expensive and time-consuming. Additionally, the model needs to be trained on a large dataset, which can also be resource-intensive.

4. Privacy Concerns: The proposed RAG model involves processing sensitive customer data, such as names and email addresses, which raises privacy concerns. It is important to ensure that the data is handled securely and complies with relevant privacy regulations.



Research

Project GraphRAG

LLM-Derived Knowledge Graphs



“GraphRAG shows substantial improvement...demonstrating intelligence or mastery that outperforms other approaches.... GraphRAG enables us to answer important classes of questions that we cannot attempt with baseline RAG alone.”



#2: Easier Development*

Easier Development

Transparent & Explicit



Customer: "I actually already fixed a couple of bugs thanks to this!"

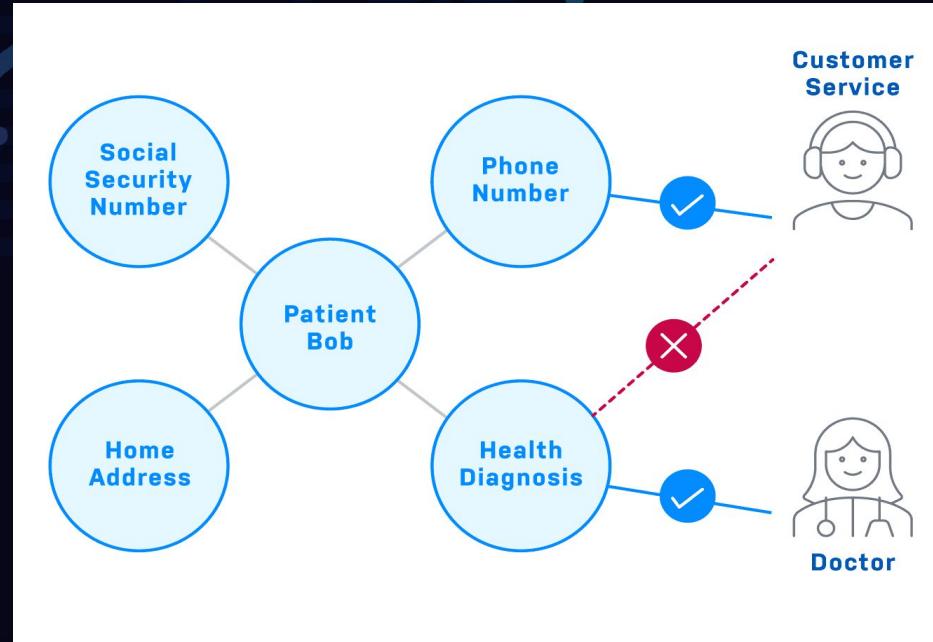
Opaque & Implicit

3.9150659e-03	2.6659777e-03	1.0298982e-03	-2.7156321e-03
1.9977870e-03	3.1204436e-03	1.2055682e-04	1.0450699e-03
-6.4308796e-04	3.0822519e-03	2.1972554e-03	5.1480172e-05
-3.7099270e-03	3.9439583e-03	6.8276987e-04	7.7137066e-04
2.3698520e-03	-7.8547641e-04	6.0383842e-04	4.6370425e-03
-1.6786088e-03	1.7417425e-03	2.4216413e-03	3.6545738e-03
-1.9871239e-03	2.9489421e-03	-1.2810023e-03	-4.9174053e-04
-3.9743204e-03	-2.7023794e-03	-3.0541950e-04	-1.5724347e-03
-2.1029566e-03	-2.1624754e-03	2.1620055e-04	-1.4000515e-03
-4.0824865e-03	4.6588355e-04	3.5028579e-03	4.8283348e-03
-2.8737928e-03	-4.5569306e-03	-7.6568732e-04	-3.3311991e-03
3.5790715e-03	4.2424244e-03	3.3478225e-03	-7.4140396e-04
1.0030111e-03	-5.2394503e-04	5.8383477e-04	-4.8430995e-03
2.6972082e-03	-4.8002079e-03	2.3011414e-03	8.0388715e-04
3.1952575e-05	-8.1621204e-04	-3.8127291e-03	-6.7428290e-04
-1.7713077e-03	-3.0159748e-03	1.7178850e-03	-1.9258332e-03
-2.4637436e-03	3.3779652e-03	2.7676420e-03	1.8853768e-03
-2.4718521e-03	-1.9754141e-03	2.6104036e-03	-2.1335895e-03
2.4405334e-03	-3.2013952e-04	3.9961869e-03	4.0419102e-03
2.0586823e-03	4.9897884e-03	4.5599132e-03	-1.0976522e-03
1.5563263e-03	3.9063310e-03	-2.9308300e-03	-4.8254002e-03
-8.7642738e-06	3.9748671e-03	5.2895391e-04	6.3330121e-04
-1.2614765e-03	-8.5018738e-04	3.7659388e-03	3.0237564e-03
4.5014662e-03	4.3258793e-03	-4.2659100e-03	4.9081761e-03
-3.9214552e-03	-2.4262110e-03	-8.1192164e-05	-4.1112076e-031



#3: Explainability and Governance

Explainability & Governance



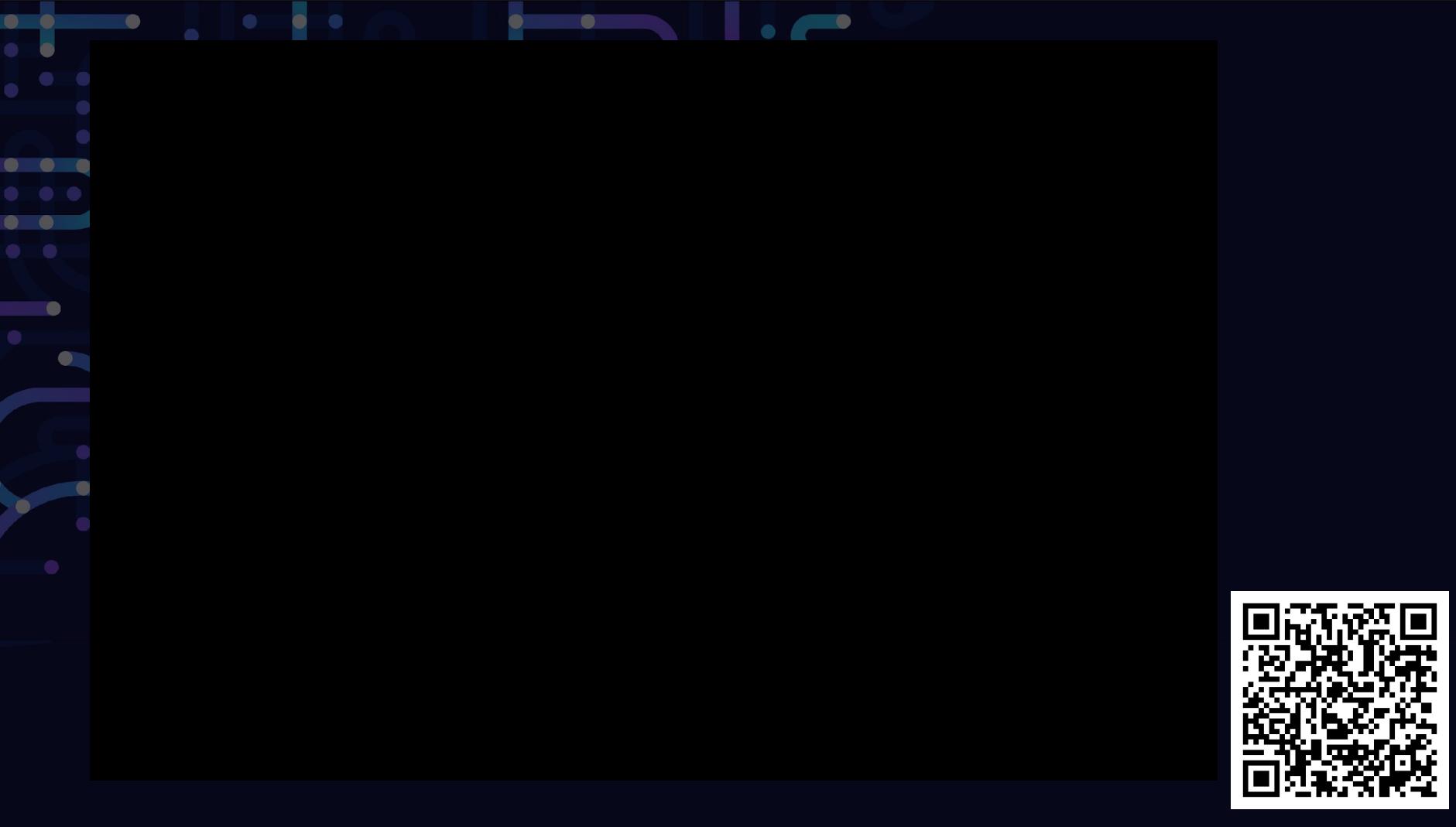
“Neo4j enables us to enforce robust information security by applying access controls at the subgraph level.”
-Eric Costantini, Chief Business Officer at Data^2.

The Benefits of GraphRAG

1. Quality & Accuracy

2. Easier Development

3. Explainability & Governance





Want to Learn More?



A screenshot of a YouTube video player. The video features a man in a light blue button-down shirt smiling. In the background, there's a whiteboard or screen with the text "Graphs for RAG" and "Das Kollegger" along with the DeepLearning.AI logo. A subtitle at the bottom of the video frame reads: "Knowledge graphs are an important technology for many large businesses,". The YouTube interface includes a play button, volume controls, a progress bar showing 0:08 / 3:47, and standard sharing and settings icons.



<https://www.deeplearning.ai/short-courses/knowledge-graphs-rag/>



GraphAcademy – Free, Self-Paced, Hands-on Online Training

<https://graphacademy.neo4j.com/>



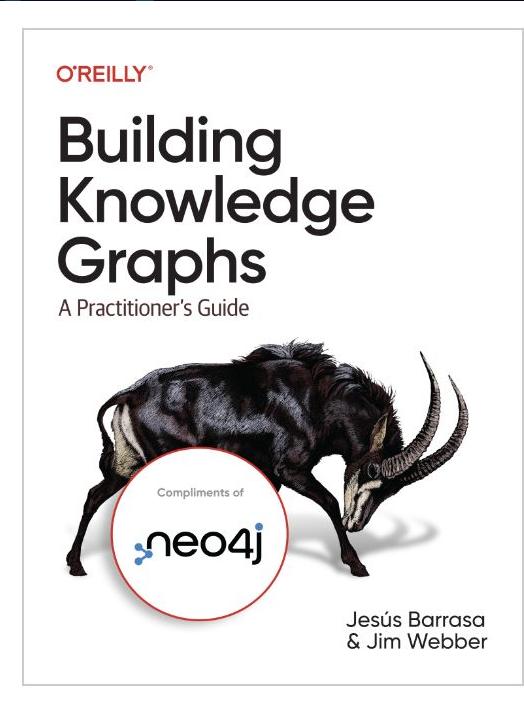
Free e-book!

Building Knowledge Graphs: A Practitioner's Guide

By Jesús Barrasa & Jim Webber

Publisher: O'Reilly

Available Formats: PDF - EN US



<https://neo4j.com/knowledge-graphs-practitioners-guide/>



The logo for Neo4j, featuring the word "neo4j" in a white, lowercase, sans-serif font.

Thank you!

Meet us at booth #907

Dr. Jesús Barrasa | Head of Solutions Architecture, Neo4j