



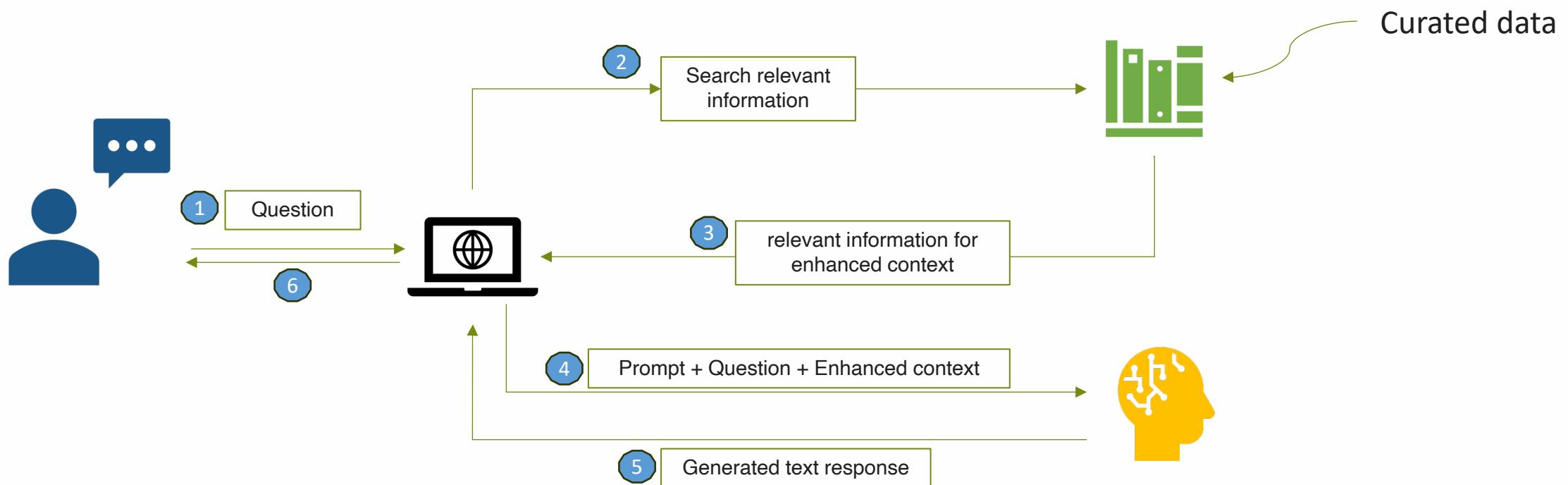
STelligence

RAG with Knowledge Graph

Nadcharin Silaphung

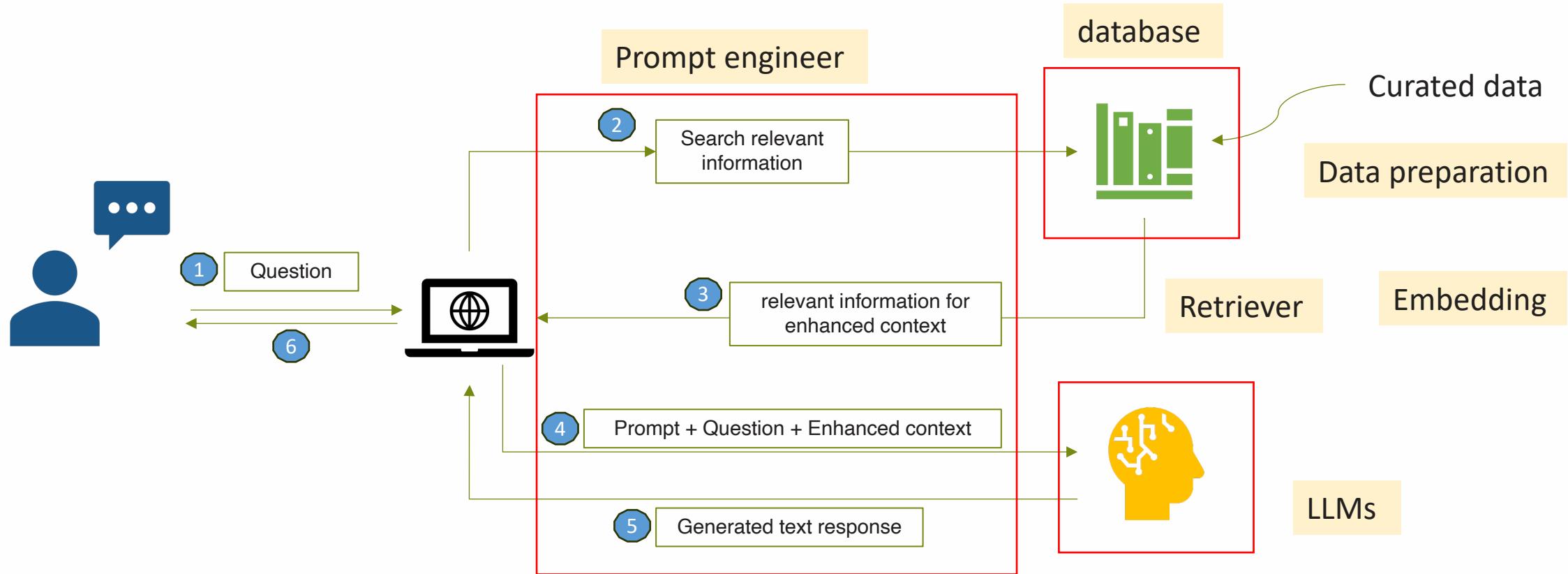
Retrieval Augmented Generation (RAG)

RAG is an approach that enhances the accuracy of LLMs by feeding relevant, real-time data outside a foundation model during the generation process.



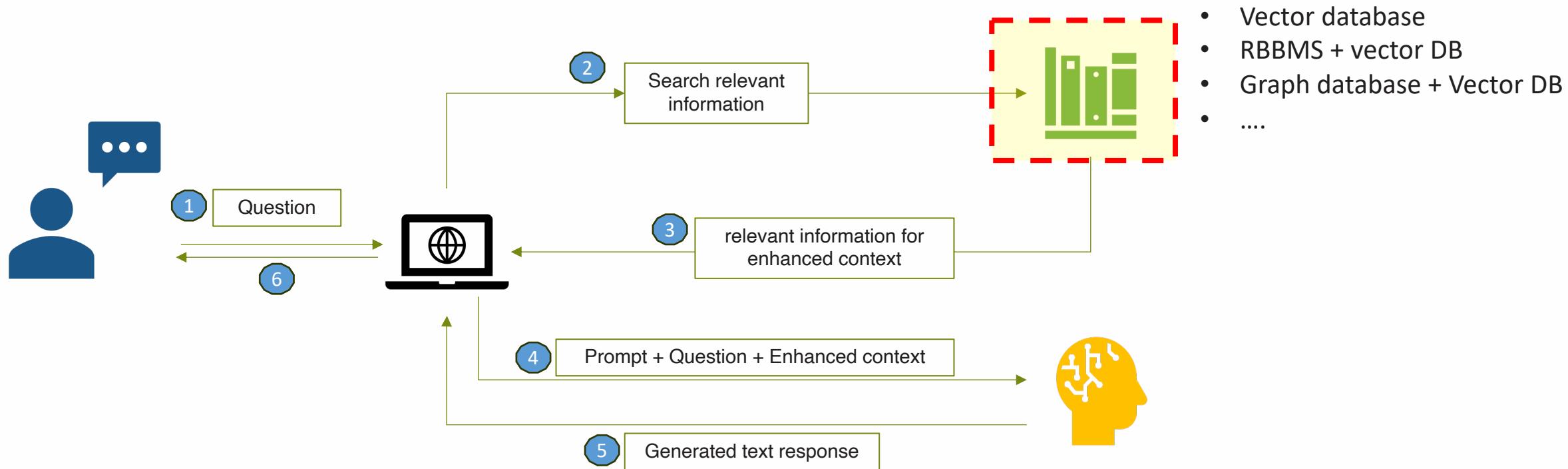
Retrieval Augmented Generation (RAG)

RAG is an approach that enhances the accuracy of LLMs by feeding relevant, real-time data outside a foundation model during the generation process.

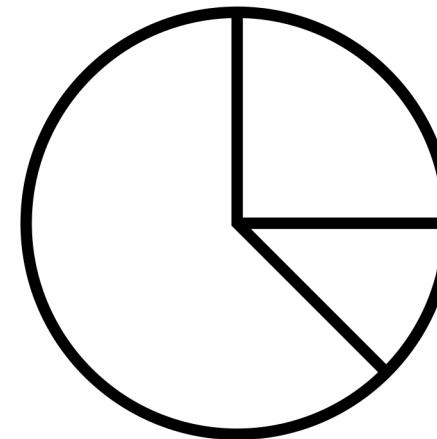
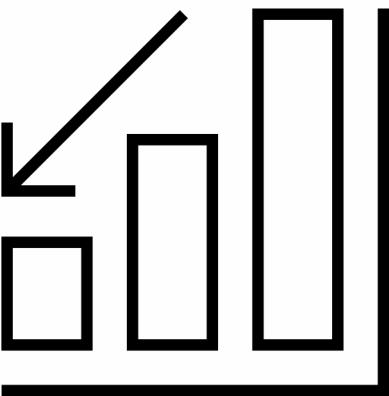
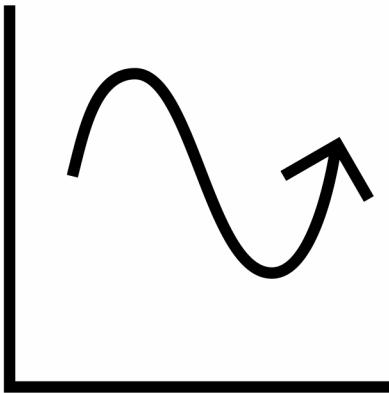


Retrieval Augmented Generation (RAG)

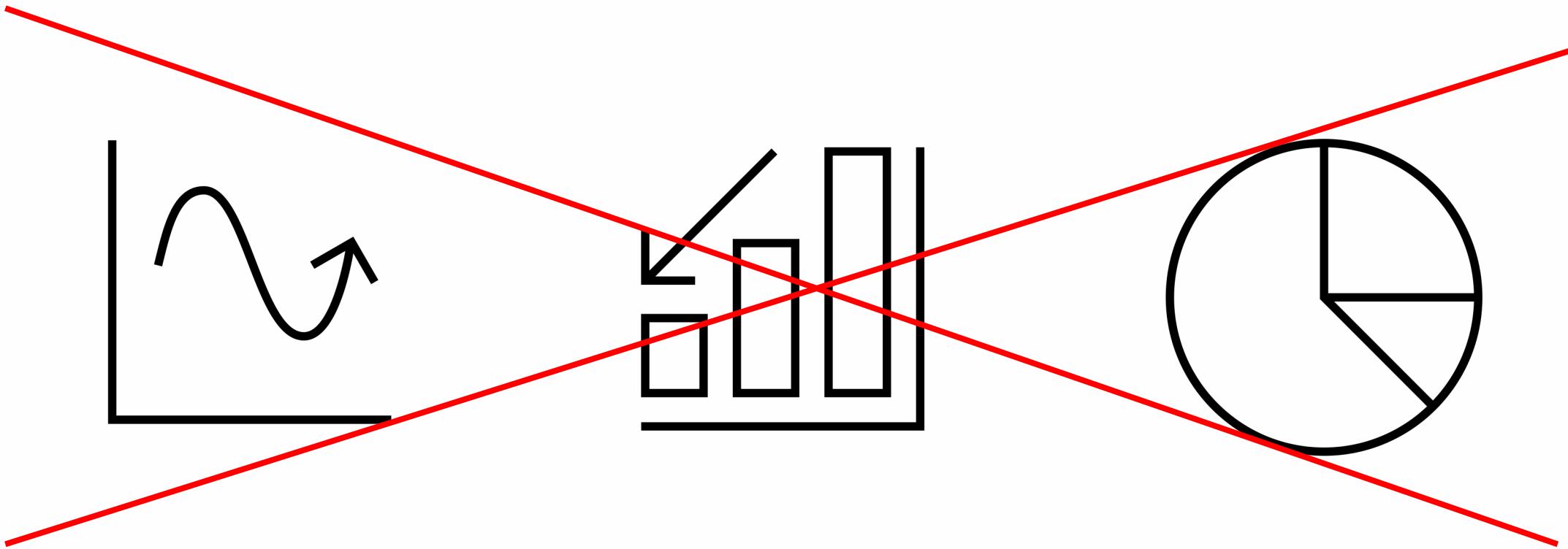
RAG is an approach that enhances the accuracy of LLMs by feeding relevant, real-time data outside a foundation model during the generation process.



What is a graph database?



What is a graph database?

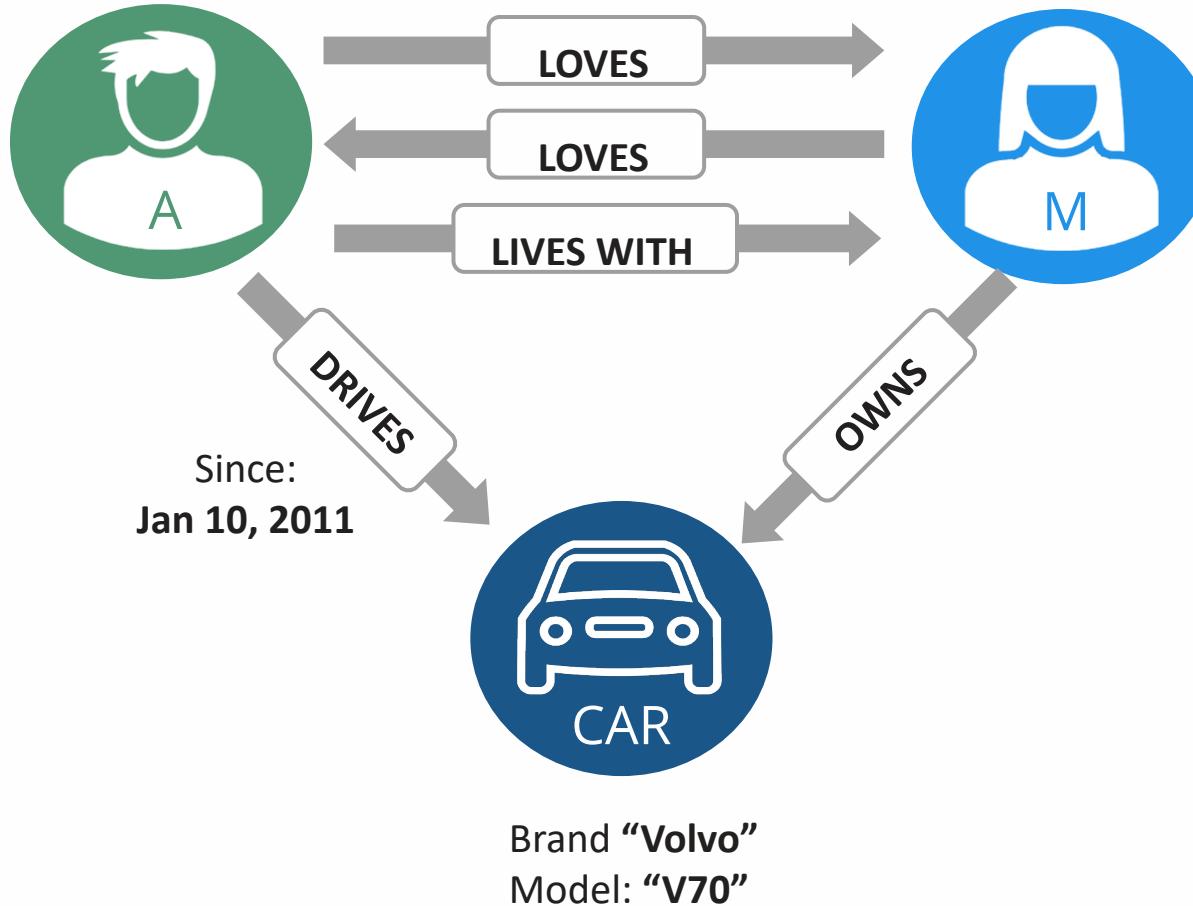


What is a graph database?

Name: "A"

Born: May 29, 1970

Twitter: "@dan"



Node

Represents an entity in the graph

Relationship

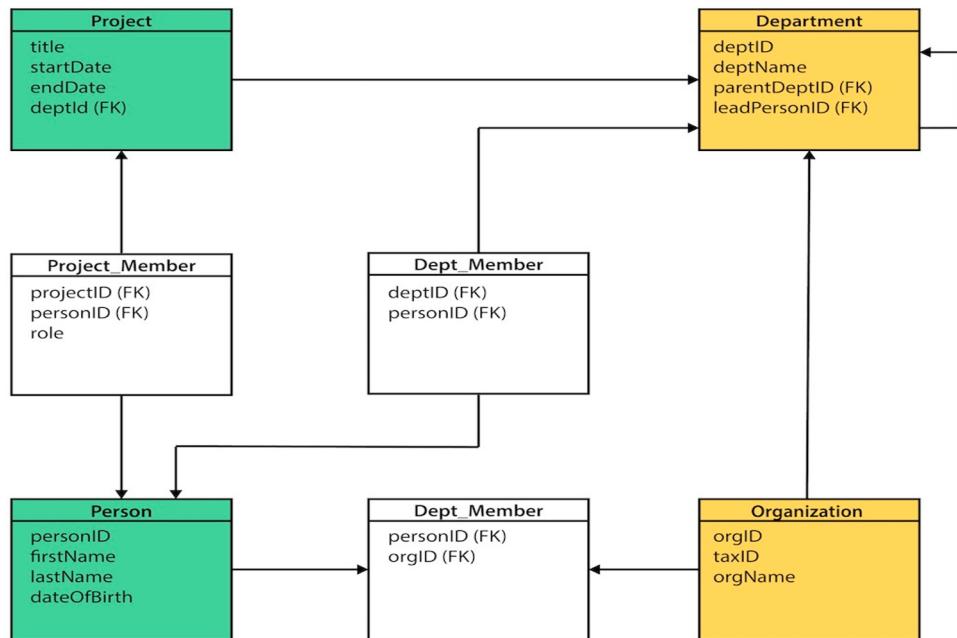
Connect nodes to each other

Property

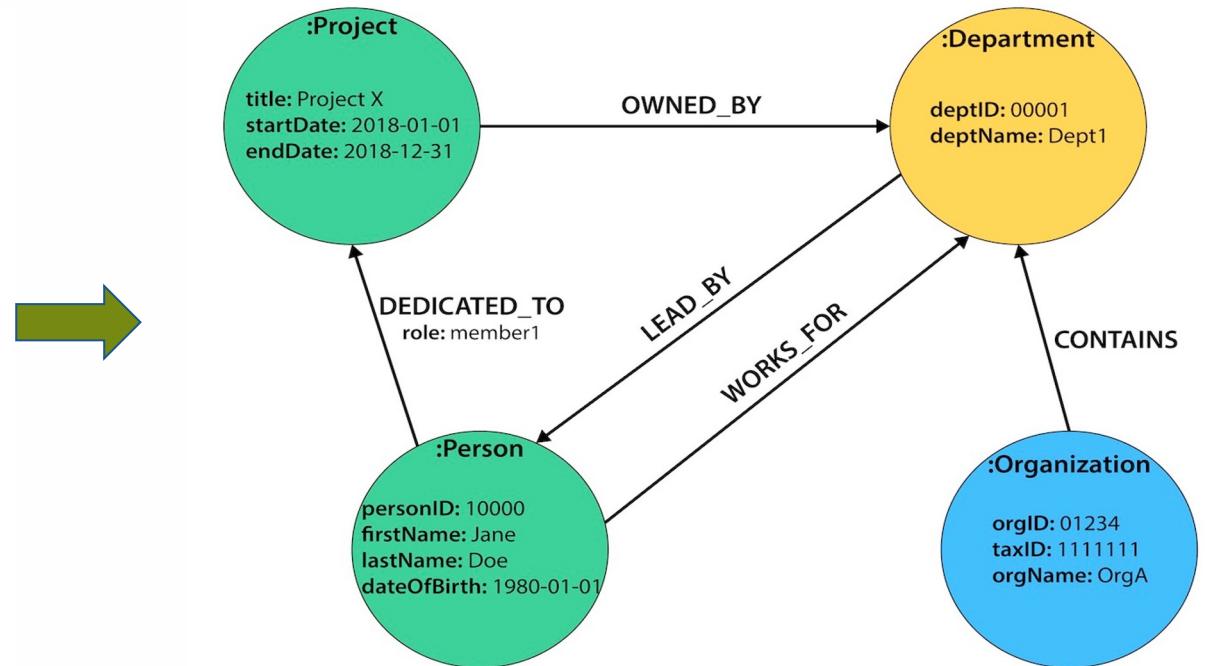
Describes a node or relationship: e.g. name, age, weight etc

Example

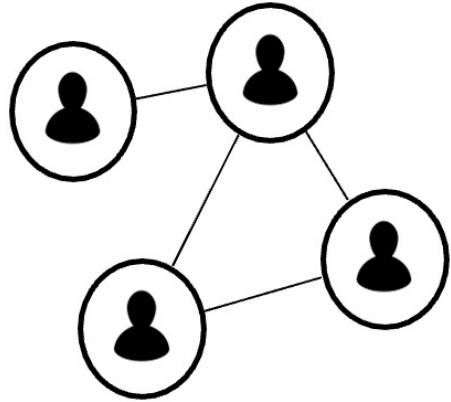
RDBMS



Graph database



Example

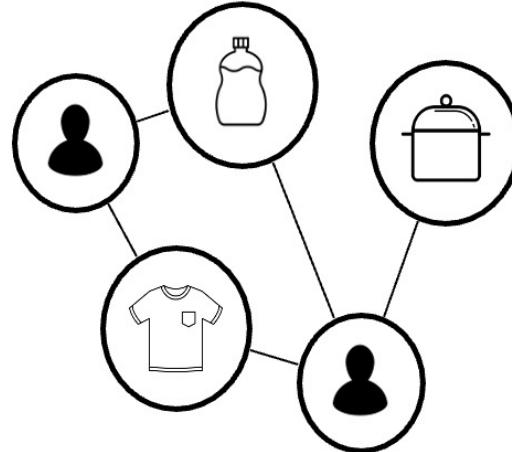


People Graph

“People you may know”

facebook

Disruptor: Facebook
Industry: Media Ad-business

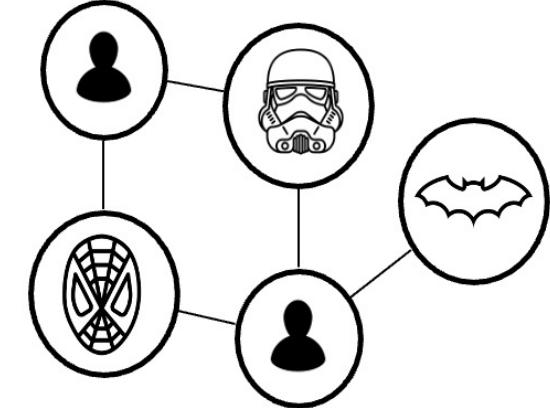


People & Products

“Other people also bought”

amazon

Disruptor: Amazon
Industry: Retail



People & Content

“You might also like”

NETFLIX

Disruptor: Netflix
Industry: Broadcasting Media

What is a Knowledge Graph

Curated data:

- The top speed of birds is 250 mph.
- The top speed of cats is 30 mph.
- The top speed of dogs is 45 mph.
- The top speed of cars is 200 mph.
- The top speed of trucks is 120mph.
- The top speed of planes is 4520 mph.

What is a Knowledge Graph

Curated data:

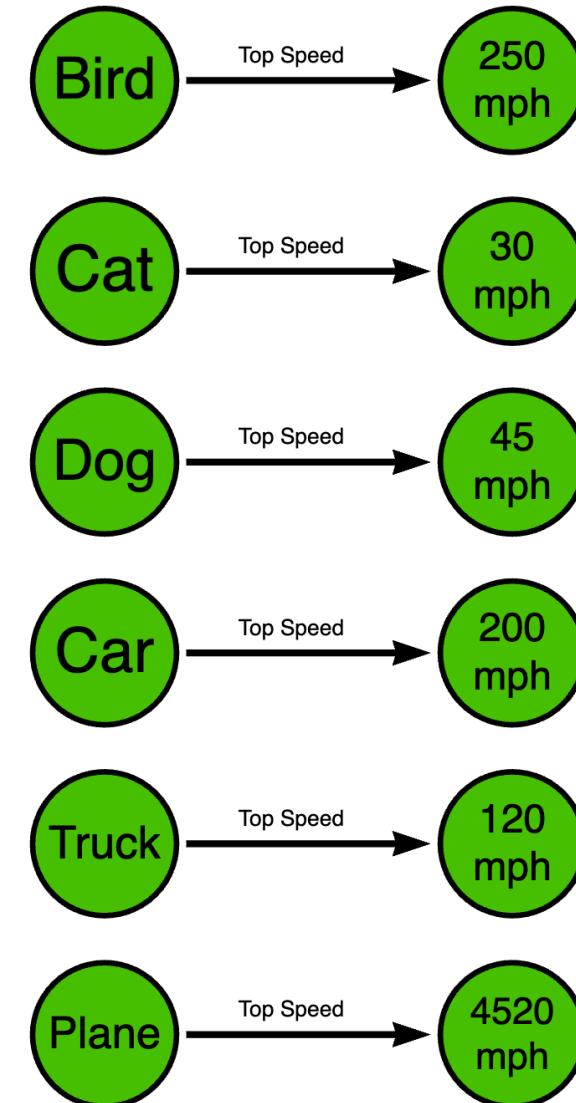
- The top speed of birds is 250 mph.
- The top speed of cats is 30 mph.
- The top speed of dogs is 45 mph.
- The top speed of cars is 200 mph.
- The top speed of trucks is 120mph.
- The top speed of planes is 4520 mph.



What is a Knowledge Graph

Curated data:

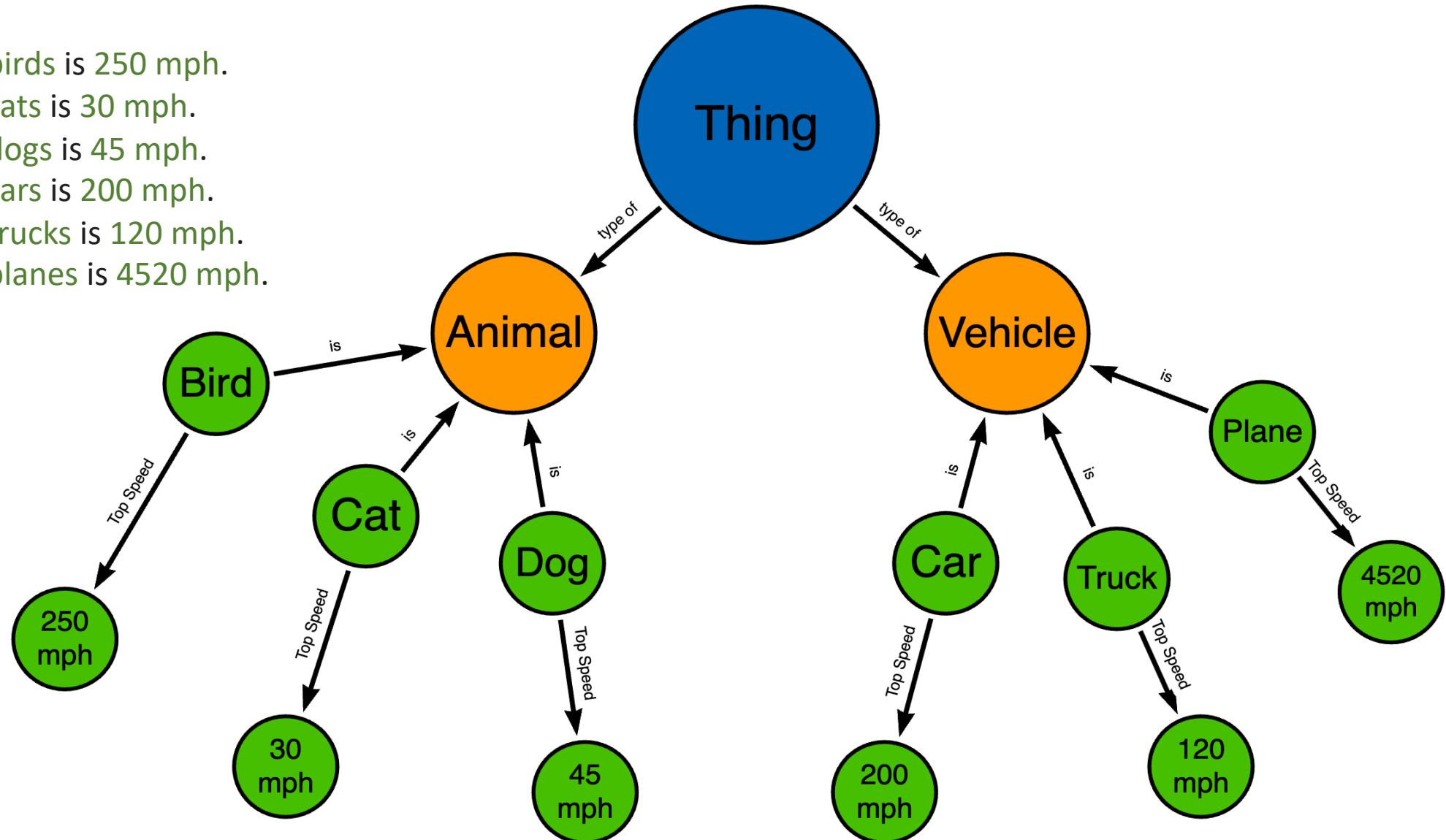
- The top speed of birds is 250 mph.
- The top speed of cats is 30 mph.
- The top speed of dogs is 45 mph.
- The top speed of cars is 200 mph.
- The top speed of trucks is 120 mph.
- The top speed of planes is 4520 mph.



What is a Knowledge Graph

Curated data:

- The top speed of birds is 250 mph.
- The top speed of cats is 30 mph.
- The top speed of dogs is 45 mph.
- The top speed of cars is 200 mph.
- The top speed of trucks is 120 mph.
- The top speed of planes is 4520 mph.



Understand and Exploit GenAI With Gartner's New Impact Radar

- Generative AI (GenAI) is rapidly evolving in a manner that offers opportunities to deliver value to clients.
- Those developing GenAI-enabled products and services must master the near-term technologies before making long-range GenAI investments.

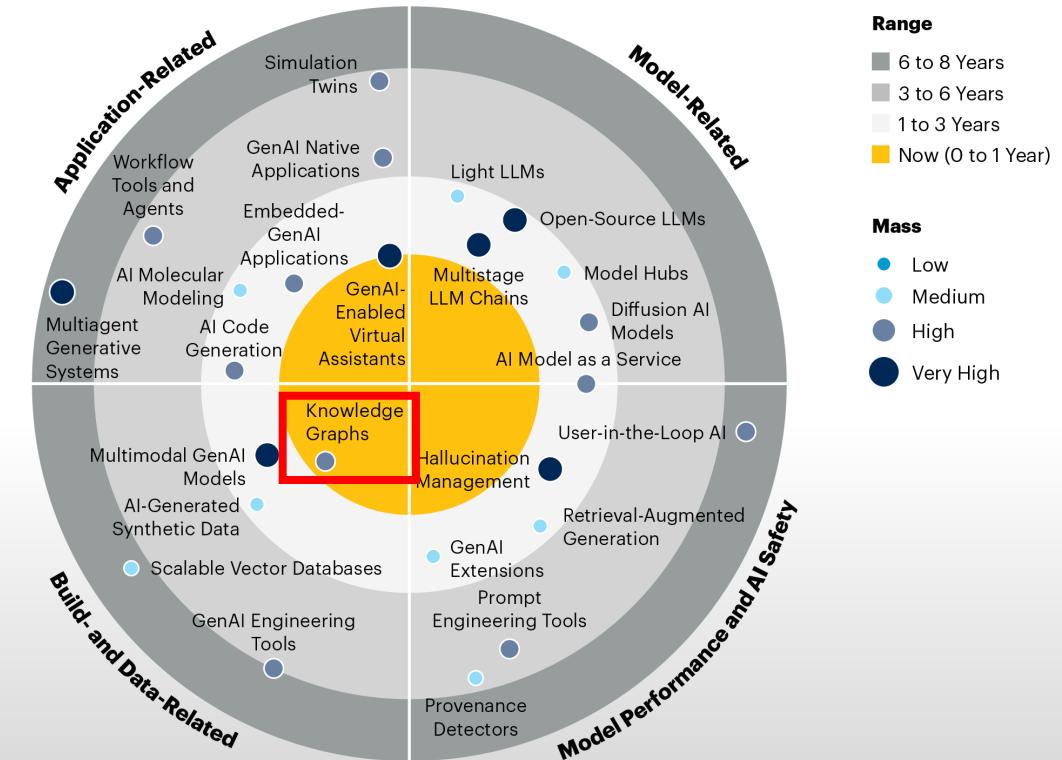
By Lori Perri | 5-minute read | December 21, 2023

Theme 3: Model build and data-related

This theme covers some of the critical steps and decisions in building and advancing a GenAI model. The following technologies and trends fall into this category:

- **Knowledge graphs (KGs)** are machine-readable data structures that represent knowledge of the physical and digital worlds, including entities and their relationships, which adhere to a graph data model.
- **Multimodal GenAI models** allow multiple types of data inputs and outputs, such as images, videos, audio, text and numerical data, within a single generative model.
- **AI-generated synthetic data** is a class of data that is often derived and extrapolated from a set of real data but is artificially generated rather than collected from real-world events.
- **Scalable vector databases** provide vector (semantic) search capability and are used in conjunction with LLMs to apply the model's ability to respond to natural language with information that is custom or specific to an enterprise or domain.
- **GenAI engineering tools** enable enterprises to operationalize models faster, balancing both governance and time to market.

Impact Radar for Generative AI



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

Gartner®

Ref: [understand-and-exploit-gen-ai-with-gartner-s-new-impact-radar](#)

LLM application (typical) VS LLM application with KG

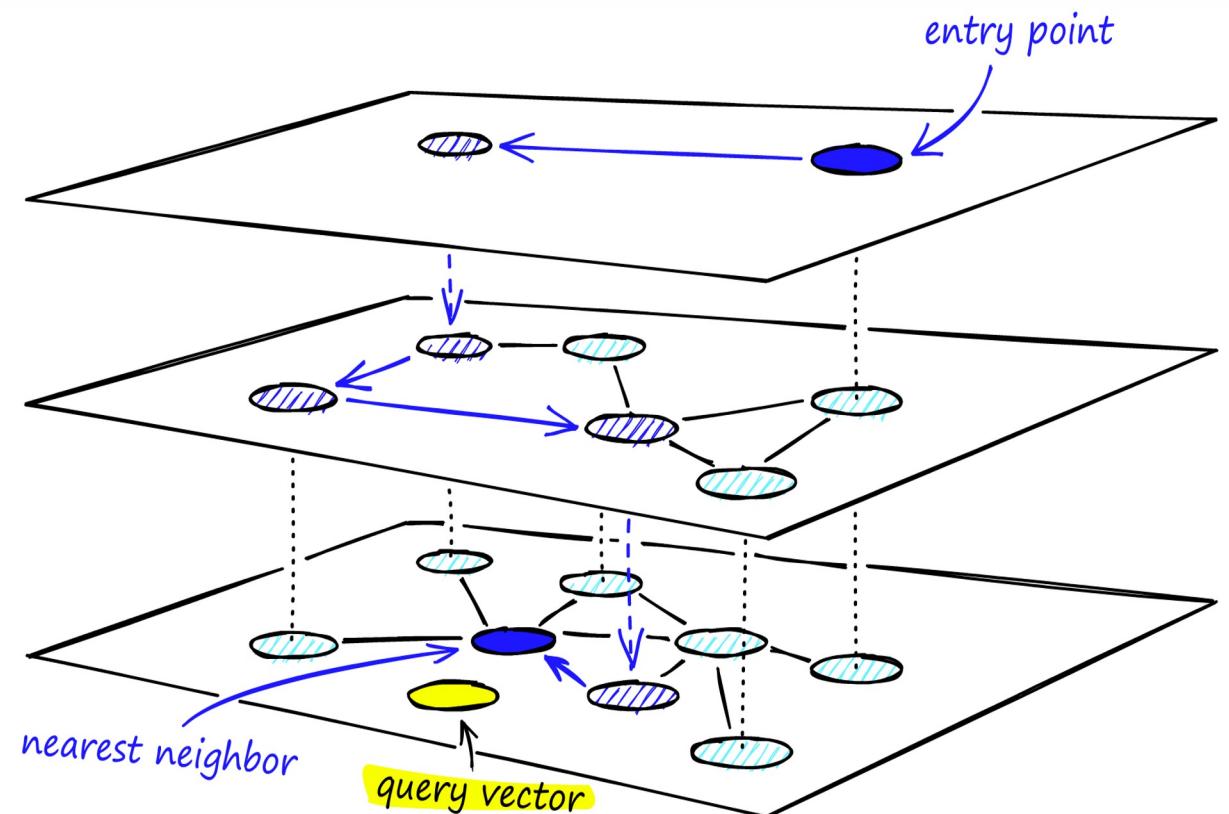
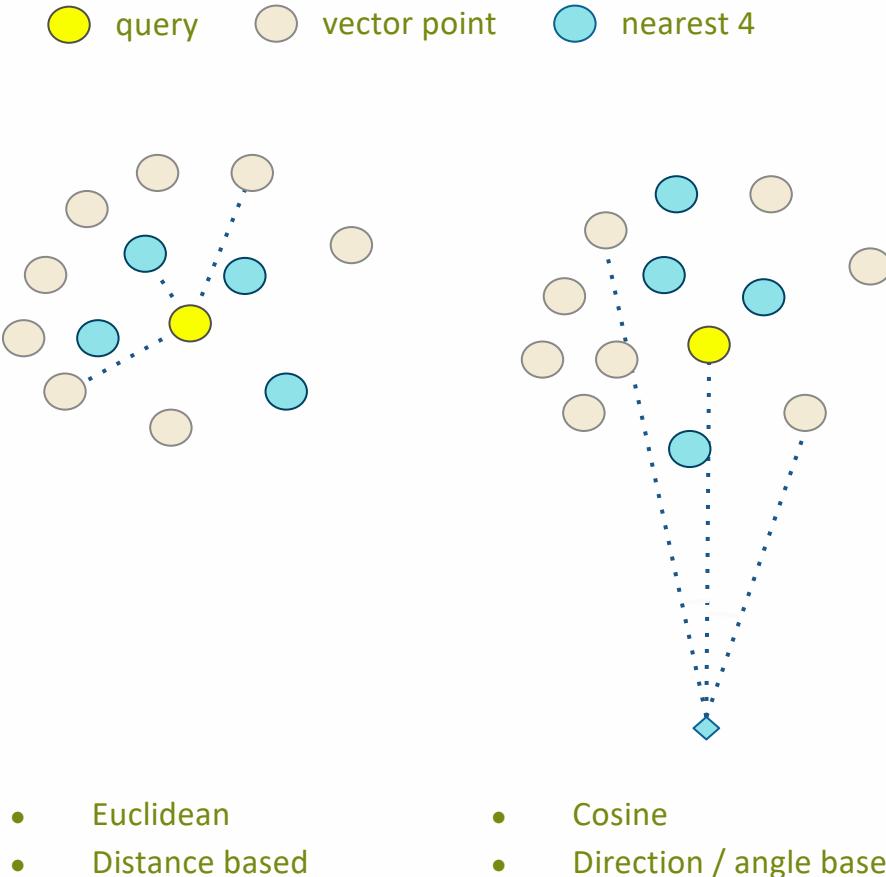


image source: [pinecone article](#)

Demo: RAG with the Harry Potter Knowledge Graph

Sources data: Information on 8 Harry Potter films from Wikipedia

Harry Potter and the Philosopher's Stone (film)

Article Talk

From Wikipedia, the free encyclopedia

Harry Potter and the Philosopher's Stone (also known as *Harry Potter and the Sorcerer's Stone*) is a 2001 fantasy film directed by [Chris Columbus](#) and produced by [David Heyman](#), from a screenplay by [Steve Kloves](#), based on the 1997 novel of the same name by J. K. Rowling. It is the first instalment in the *Harry Potter* film series. The film stars [Daniel Radcliffe](#) as Harry Potter, with [Rupert Grint](#) as Ron Weasley, and [Emma Watson](#) as Hermione Granger. Its story follows Harry's first year at [Hogwarts School of Witchcraft and Wizardry](#) as he discovers that he is a famous wizard and begins his formal wizarding education.

Warner Bros. Pictures bought the film rights to the book in 1999 for a reported £1 million (\$1.65 million). Production began in the United Kingdom in 2000, with Columbus being chosen to helm the film from a

Harry Potter and the Chamber of Secrets (film)

Article Talk

From Wikipedia, the free encyclopedia

Harry Potter and the Chamber of Secrets is a 2002 fantasy film directed by [Chris Columbus](#) from a screenplay by [Steve Kloves](#), based on the 1998 novel of the same name by J. K. Rowling. Produced by [David Heyman](#), it is the sequel to *Harry Potter and the Philosopher's Stone* (2001) and the second instalment in the *Harry Potter* film series. The film stars [Daniel Radcliffe](#) as Harry Potter, with [Rupert Grint](#) and [Emma Watson](#) as his best friends [Ron Weasley](#) and [Hermione Granger](#) respectively. The story follows Harry's second year at [Hogwarts School of Witchcraft and Wizardry](#), where the Heir of [Salazar Slytherin](#) opens the *Chamber of Secrets*, unleashing a monster that petrifies the school's students.

The film was released in the United Kingdom and the United States on 15 November 2002, by [Warner Bros. Pictures](#). Critics praised its darker plot, sets, performances (especially Branagh, Coltrane and Isaacs), and a story appropriate for a young audience, and it became a critical and commercial success, grossing \$926 million worldwide and becoming the second-highest-grossing film of 2002. The film was nominated for many awards, including the [BAFTA Award for Best Production Design](#), [Best Sound](#), and [Best Special Visual Effects](#). It was followed by *Harry Potter and the Prisoner of Azkaban* (2004).

84 languages

Read Edit View history Tools

+



Harry Potter and the Prisoner of Azkaban (film)

Article Talk

From Wikipedia, the free encyclopedia

Harry Potter and the Prisoner of Azkaban is a 2004 fantasy film directed by [Alfonso Cuarón](#) from a screenplay by [Steve Kloves](#), based on the 1999 novel of the same name by J. K. Rowling. It is the sequel to *Harry Potter and the Chamber of Secrets* (2002) and the third instalment in the *Harry Potter* film series. The film stars [Daniel Radcliffe](#) as Harry Potter, alongside [Rupert Grint](#) and [Emma Watson](#) as Harry's best friends [Ron Weasley](#) and [Hermione Granger](#) respectively. The film follows Harry's third year at [Hogwarts](#) and his quest to uncover the truth about his past, including the connection recently-escaped [Azkaban](#) prisoner [Sirius Black](#) has to Harry and his deceased parents.

With this film, the *Harry Potter* series switched to a longer eighteen-month production cycle. Cuarón was selected as director from a list that included [Callie Khouri](#) and [Kenneth Branagh](#). The cast of previous instalments returned for the film, with the additions of [Gary Oldman](#), [David Thewlis](#), and [Emma Thompson](#), among others. It was the first appearance of [Michael Gambon](#) as Professor [Albus Dumbledore](#), due to [Richard Harris](#)'s death in 2002. Principal photography began in February 2003 at [Leavesden Film Studios](#). It was the first in the series to extensively use real-life locations, with sets built in Scotland and scenes shot in London. Filming concluded in November 2003.

The film was released on 31 May 2004 in the United Kingdom, and on 4 June 2004 in North America. It was the first *Harry Potter* film to use IMAX Technology and released into IMAX theatres. *Prisoner of Azkaban* grossed a total of \$798 million worldwide, making it the second-highest-grossing film of 2004, behind *Shrek 2*. Critics praised Cuarón's direction and the lead actors' performances. It is credited for marking a notable change in the franchise's tone and directing style and is often regarded by critics and fans alike as the best *Harry Potter* film. It was nominated for two Academy Awards, Best Original Music Score and Best Visual Effects, at the 77th Academy Awards in 2004. It was followed by *Harry Potter and the Goblet of Fire* in 2005.

77 languages

Read Edit View history Tools

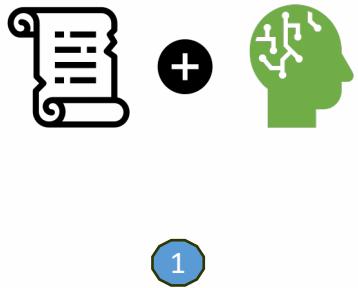
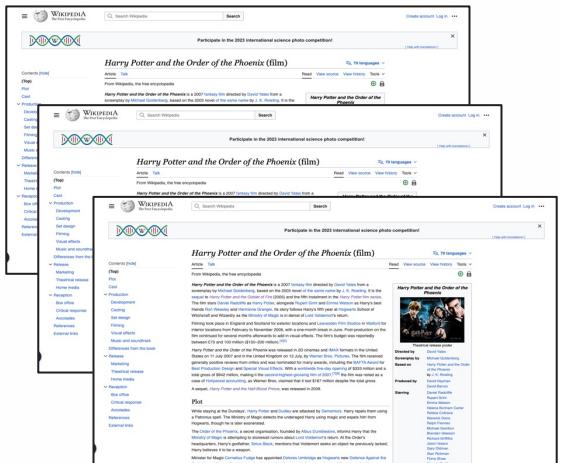
Harry Potter and the Prisoner of Azkaban



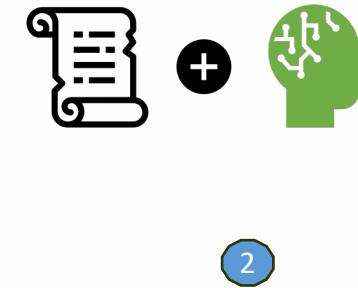
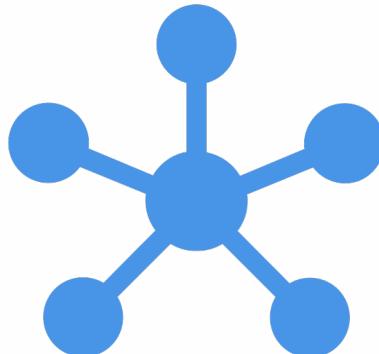
Theatrical release poster

Directed by	Alfonso Cuarón
Screenplay by	Steve Kloves
Based on	<i>Harry Potter and the Prisoner of Azkaban</i> by J. K. Rowling
Produced by	David Heyman Chris Columbus Mark Radcliffe
Starring	Daniel Radcliffe Rupert Grint Emma Watson Robbie Coltrane Michael Gambon Richard Griffiths

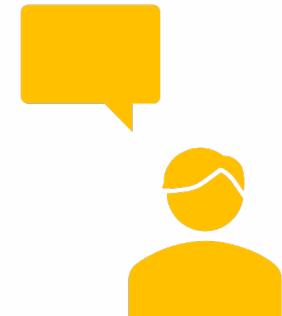
Demo: RAG with the Harry Potter Knowledge Graph



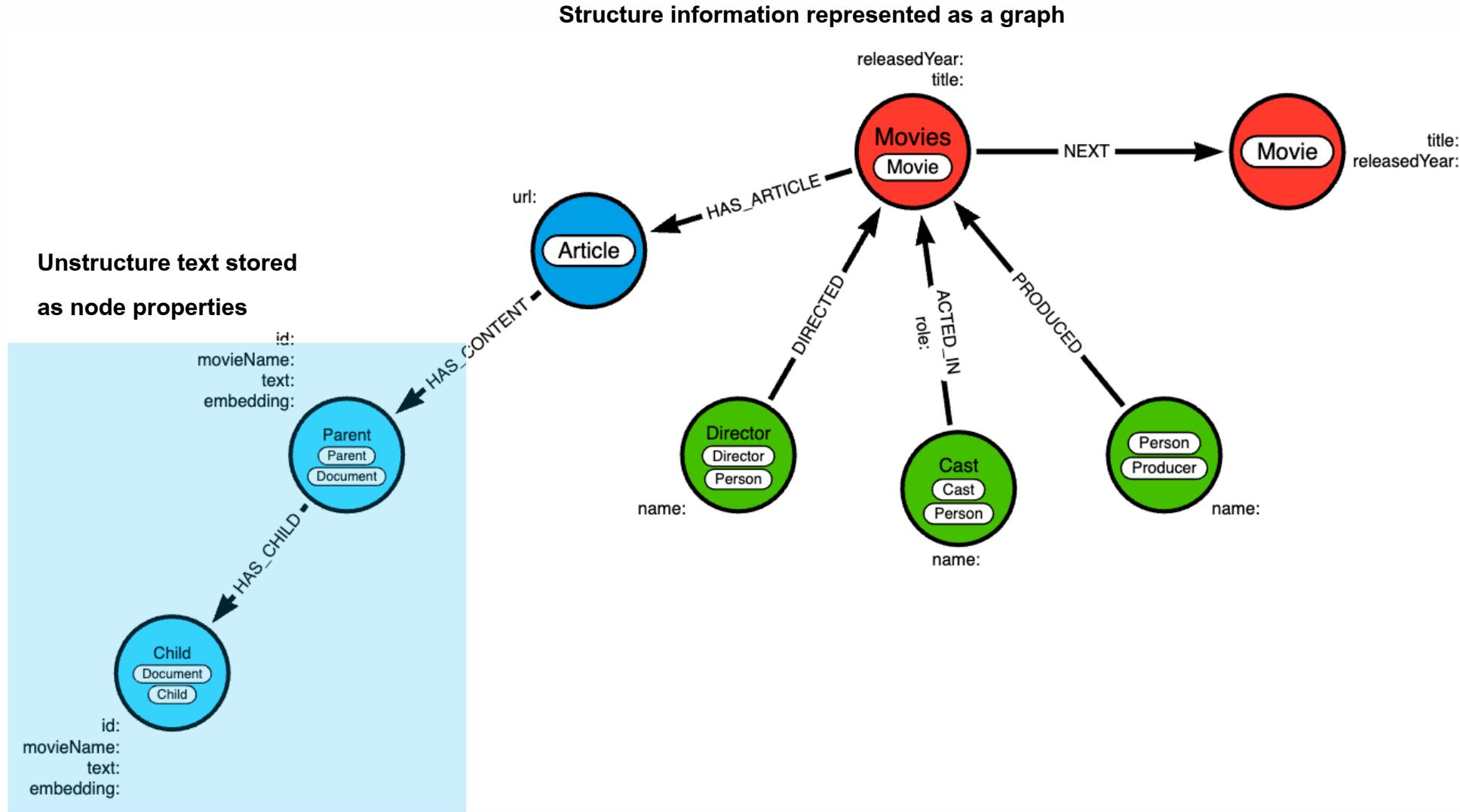
Data processing



serving



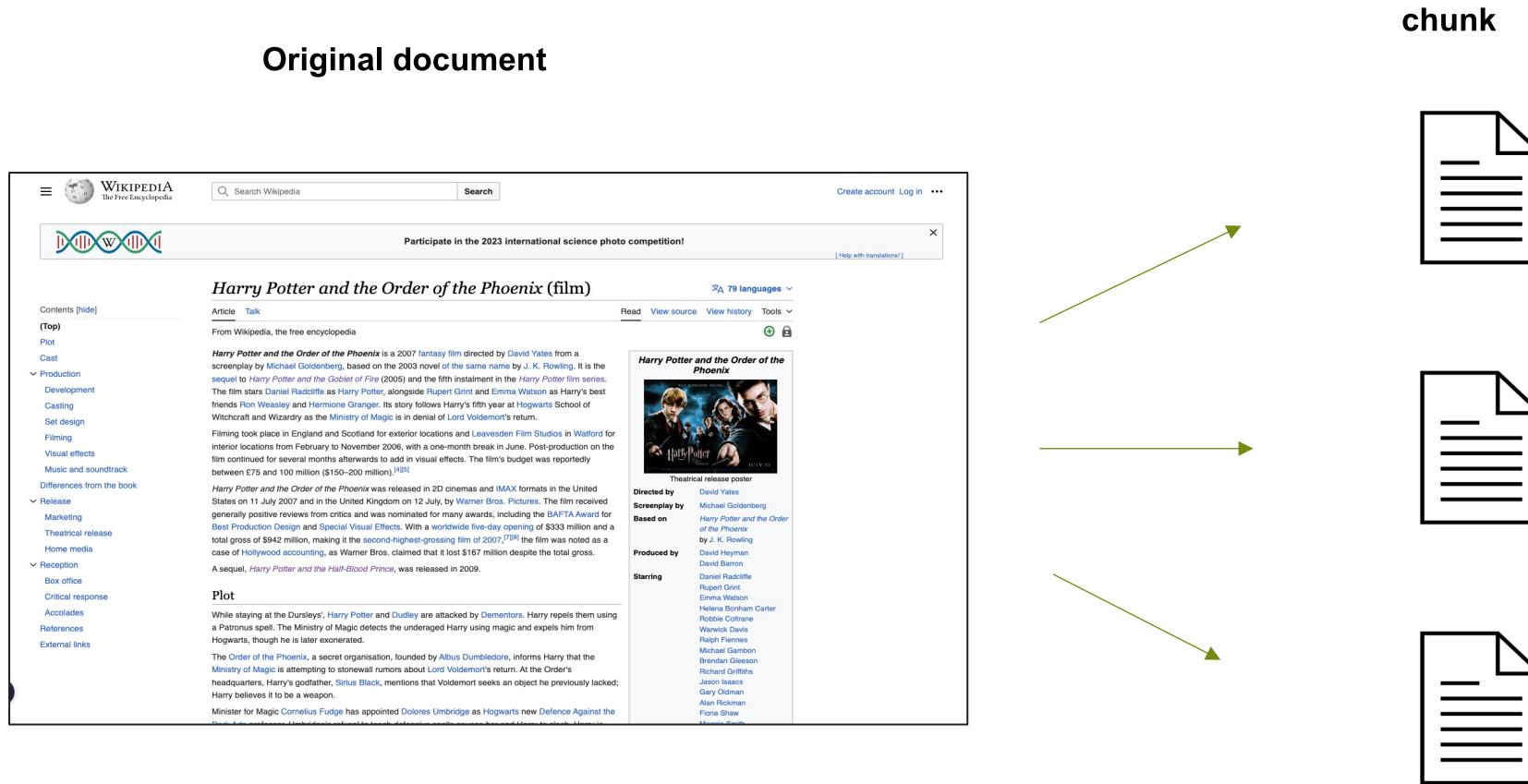
Demo: RAG with the Harry Potter Knowledge Graph



Demo: RAG with the Harry Potter Knowledge Graph

Data chunking |

Traditional method where the exact data indexed is the data retrieved.

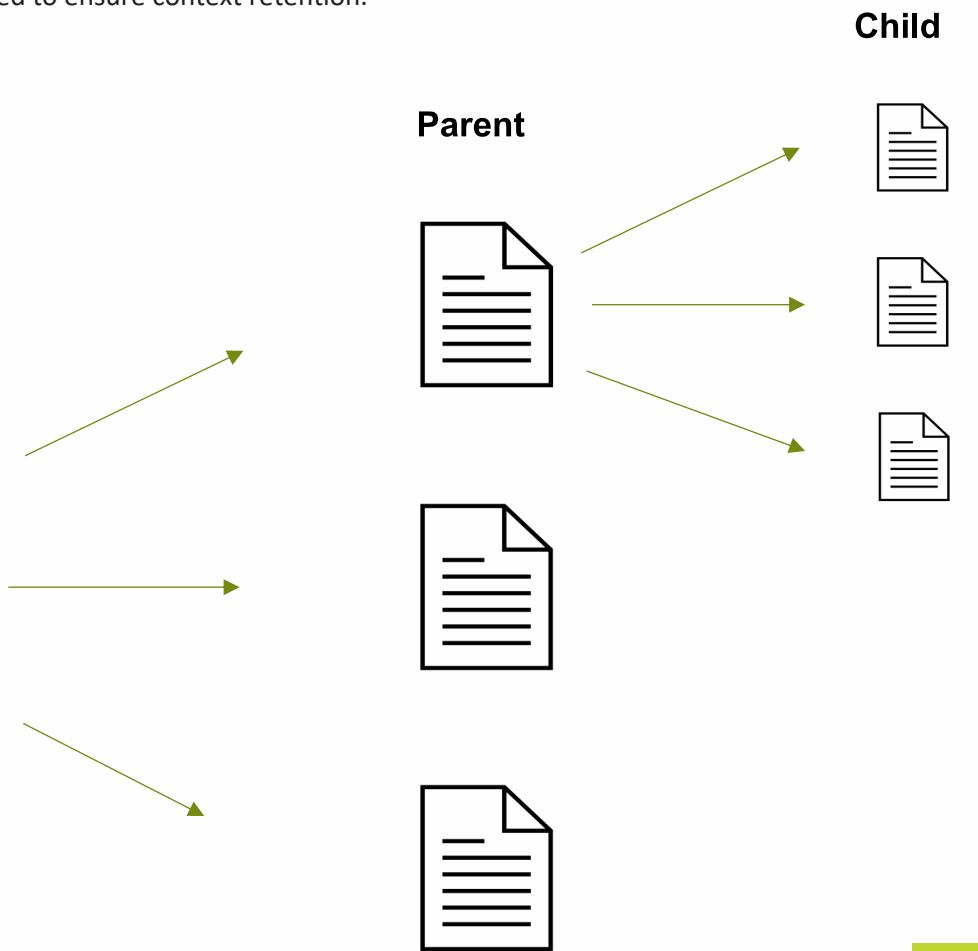
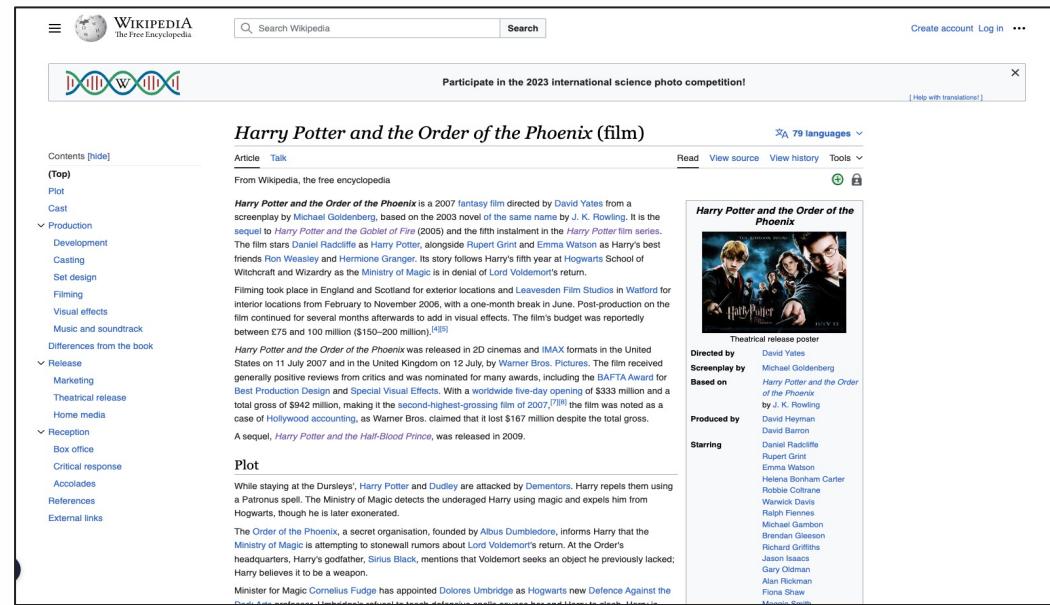


Demo: RAG with the Harry Potter Knowledge Graph

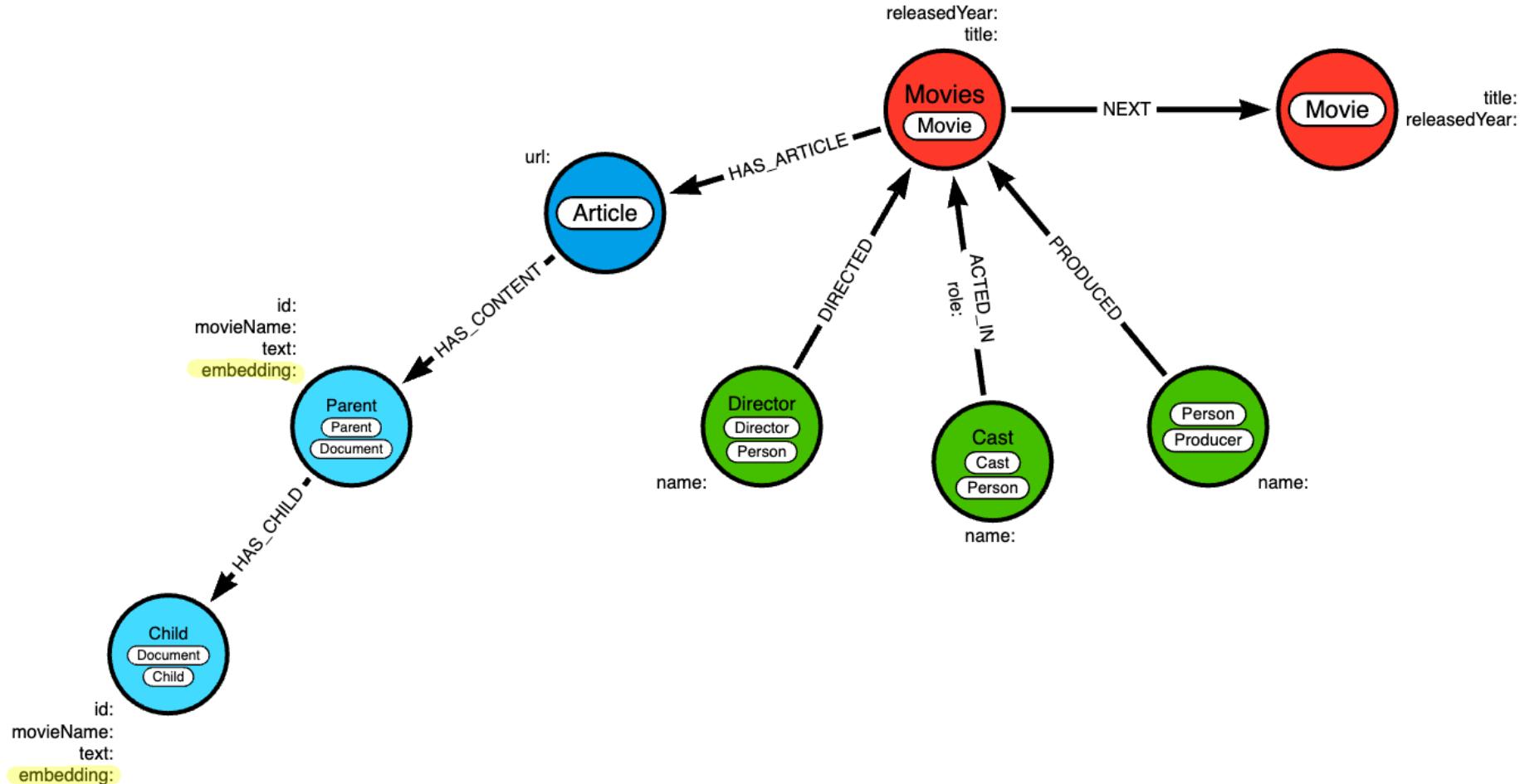
Data chunking ||

- Instead of indexing entire documents, data is divided into smaller chunks, referred to as Parent and Child documents.
- Child documents are indexed for better representation of specific concepts, while parent documents are retrieved to ensure context retention.

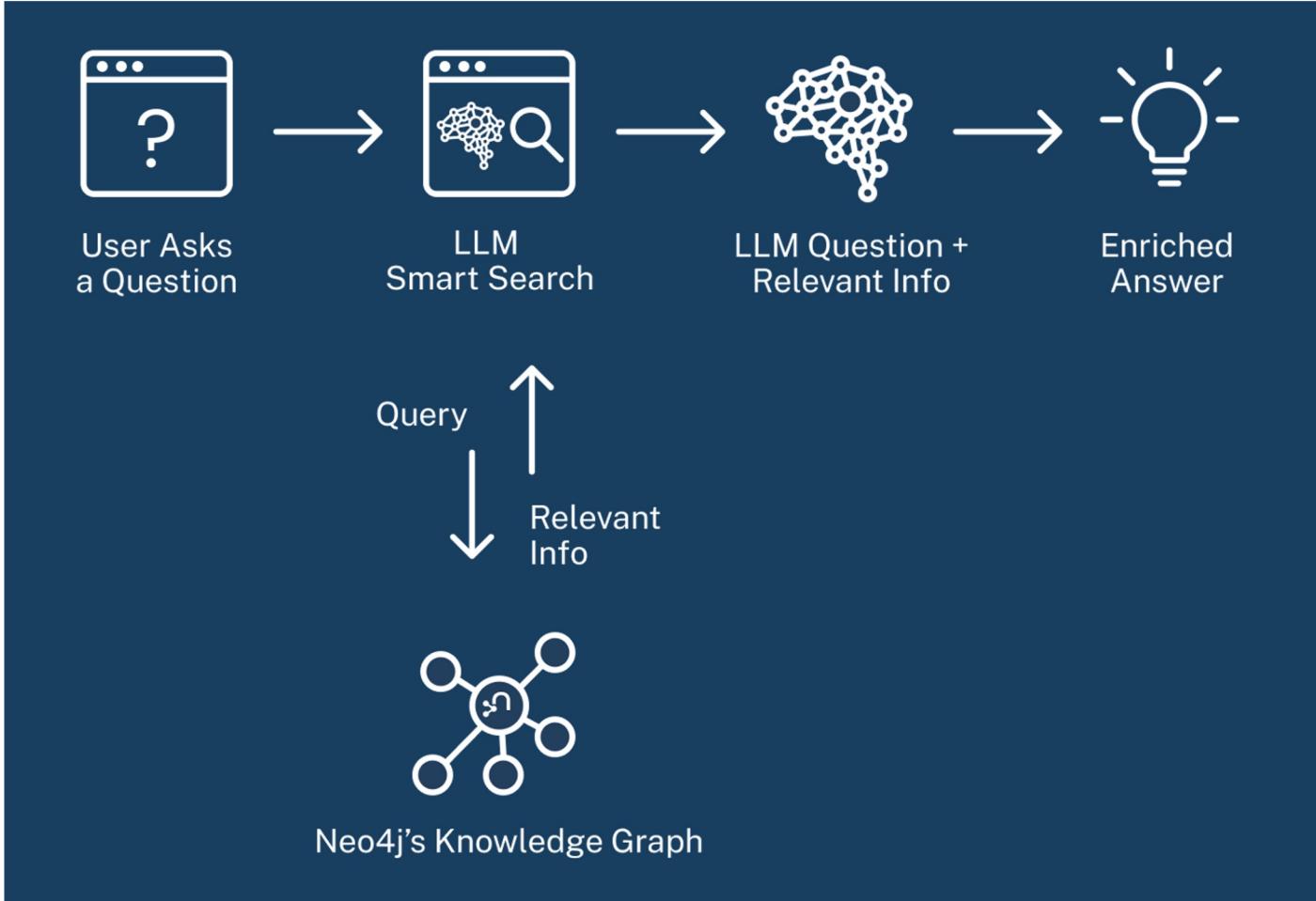
Original document



Demo: RAG with the Harry Potter Knowledge Graph



Retrieving data from KG Agent



1. Generate **embedding** for text OR extract context with prompting
2. For a given question, retrieve text based on embedding **similarity** or plain cypher queries in a knowledge store
3. Combine retrieved text and instruction to be the **prompt** for LLM to generate answer.

Making LLM application Smarter

1. The Basics

- **Knowledge Graphs:** Think of a giant web that connects all kinds of information, like who knows who, and what is related to what. It's like a map for AI to understand relationships.
- **Embeddings:** Imagine teaching AI to understand similarities between things by feeling their "vibes." It's like turning concepts into a special code that AI can easily compare.

2. Working Together

- **Mixing the Codes:** We combine the map (knowledge graphs) and vibes (embeddings) so AI can get a deeper understanding of everything.
- **Finding Answers:** First, AI uses the map to see the big picture, then uses vibes to quickly find related stuff.

3. Why It's Cool

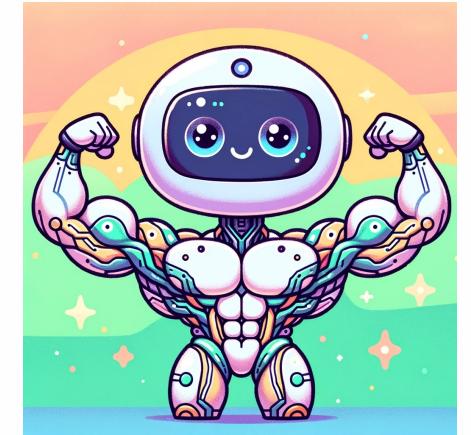
- **Easy to Explain:** Because AI uses a map, we can follow its path to understand how it makes decisions.
- **Getting Smarter:** The more AI learns, the better it gets at adding new information to the map and understanding new vibes.

4. What We Get

- A super smart AI that can understand complex stuff by combining the best of both worlds: detailed maps and intuitive vibes, making it great at answering tough questions.

Key takeaway

Knowledge graph + Vector Embedding =



Resources & Materials

- Graph Academy : <https://graphacademy.neo4j.com>
- Resource : <https://neo4j.com/resources>
- Book on knowledge graph: <https://neo4j.com/books/knowledge-graphs-practitioners-guide>

?

Q & A

