



Language Models and Structured Knowledge in AI

12.09.2024

Dr. Gerrit Großmann, Islam Mesabah

github.com/gerritgr/LLMSummerSchool



dfki
ai
Deutsches Forschungszentrum
für Künstliche Intelligenz
German Research Center for
Artificial Intelligence

What to expect

Theory (30 min)



General introduction into multiple ideas



Details of every neuro-symbolic LLM concept, LLM agents, prompt engineering

Hands-on (2 hours)



Implementation of a KG-based RAG system based on Mistral and neo4j



Revising every software library for combining LLMs with structured knowledge

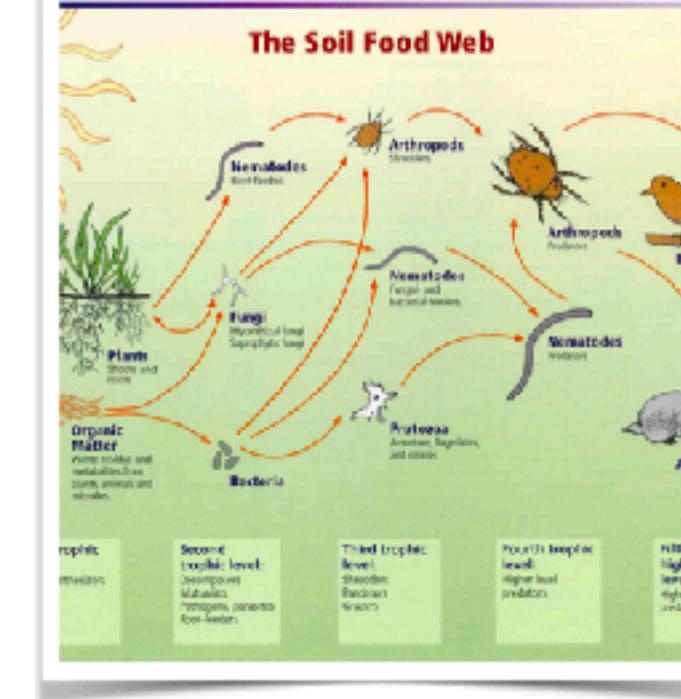
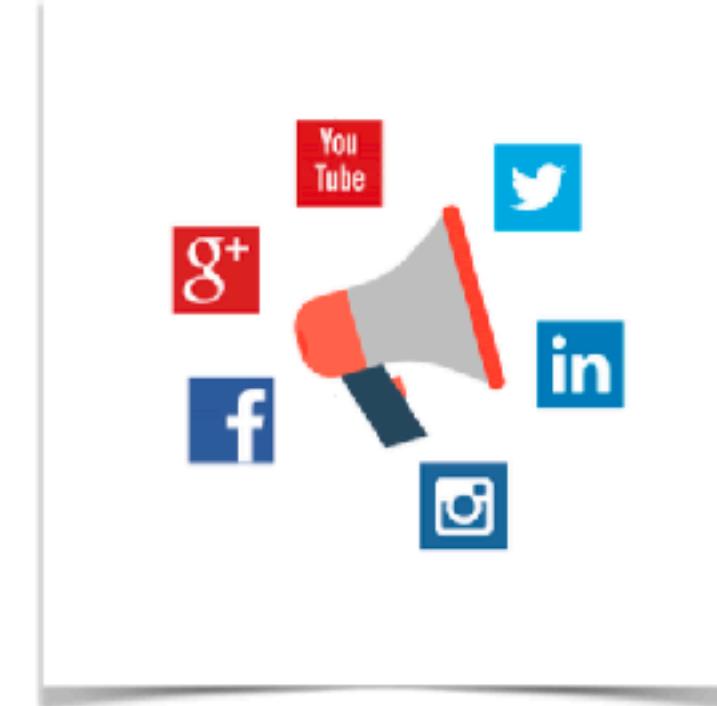
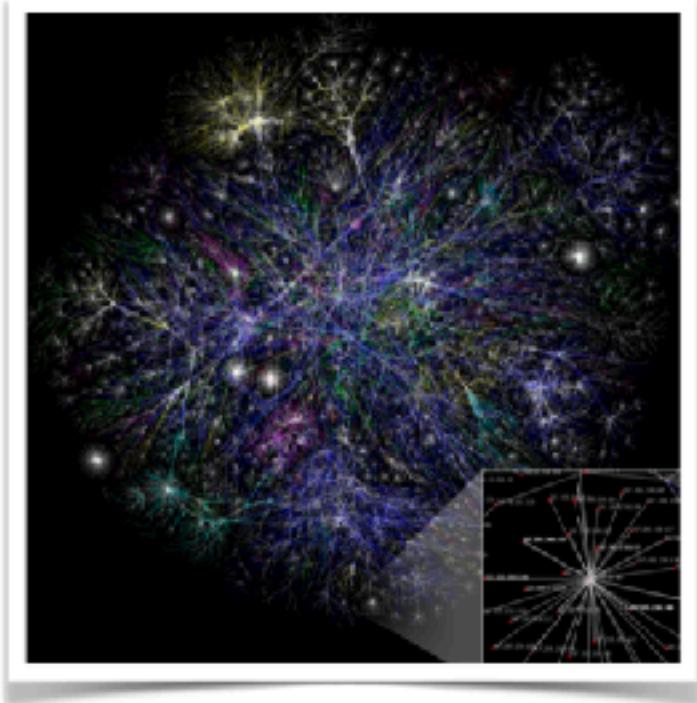
You know this.

Language Models and Structured Knowledge in AI

... not so obvious.

- Databases & tabular data
- XML and JSON files
- Programm code & rule-based models
- Knowledge graphs
- Labeled files

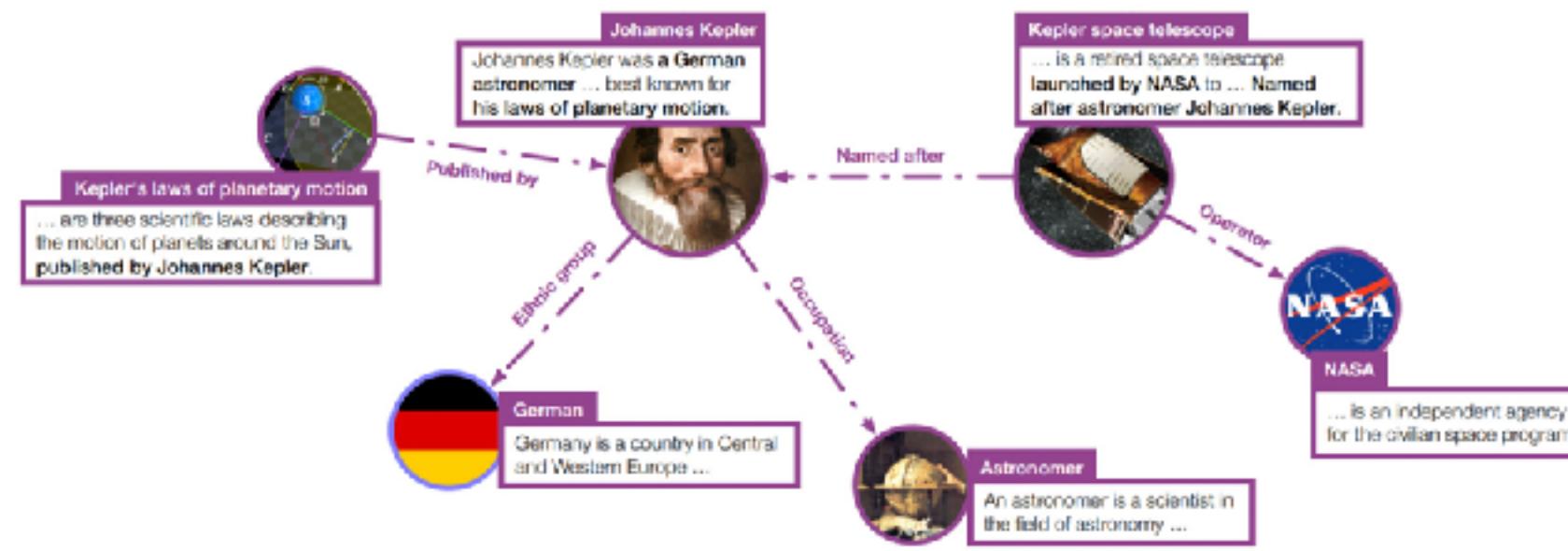
Why Graphs?



In many ways, graphs are the main modality of data we receive from nature.

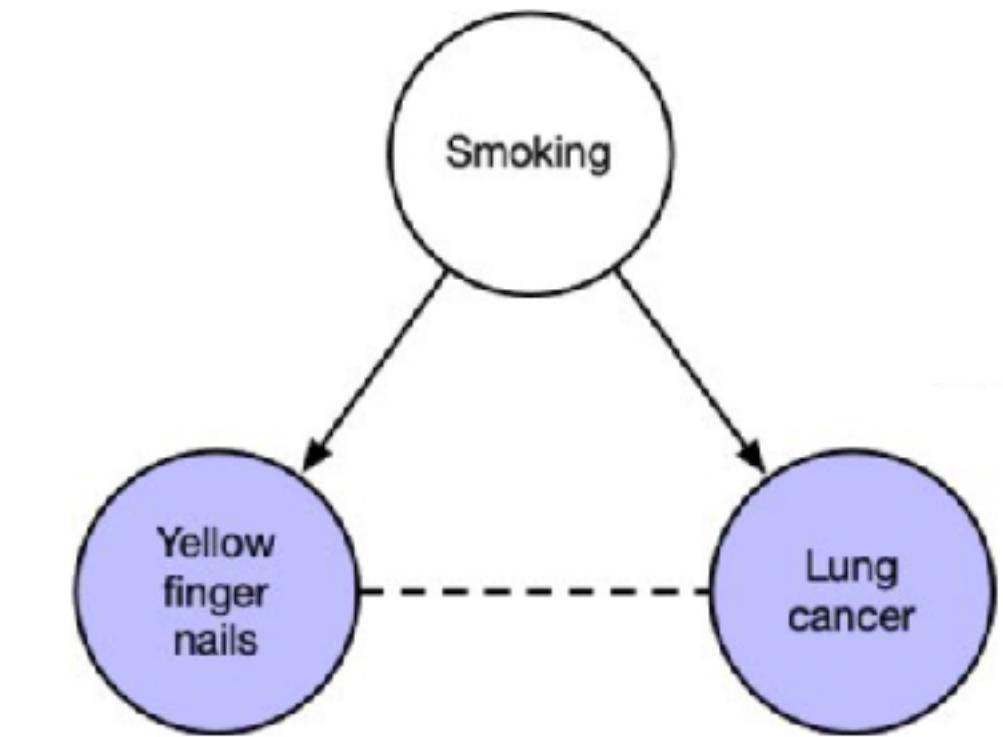
- Petar Veličković

Why Graphs?



<https://deepgraphlearning.github.io/project/wikidata5m>

Knowledge Graph



An introduction to causal reasoning in health analytics (Zhang et al)

Causal Graph



<https://towardsdatascience.com/visualizing-protein-networks-in-python-58a9b51be9d5>

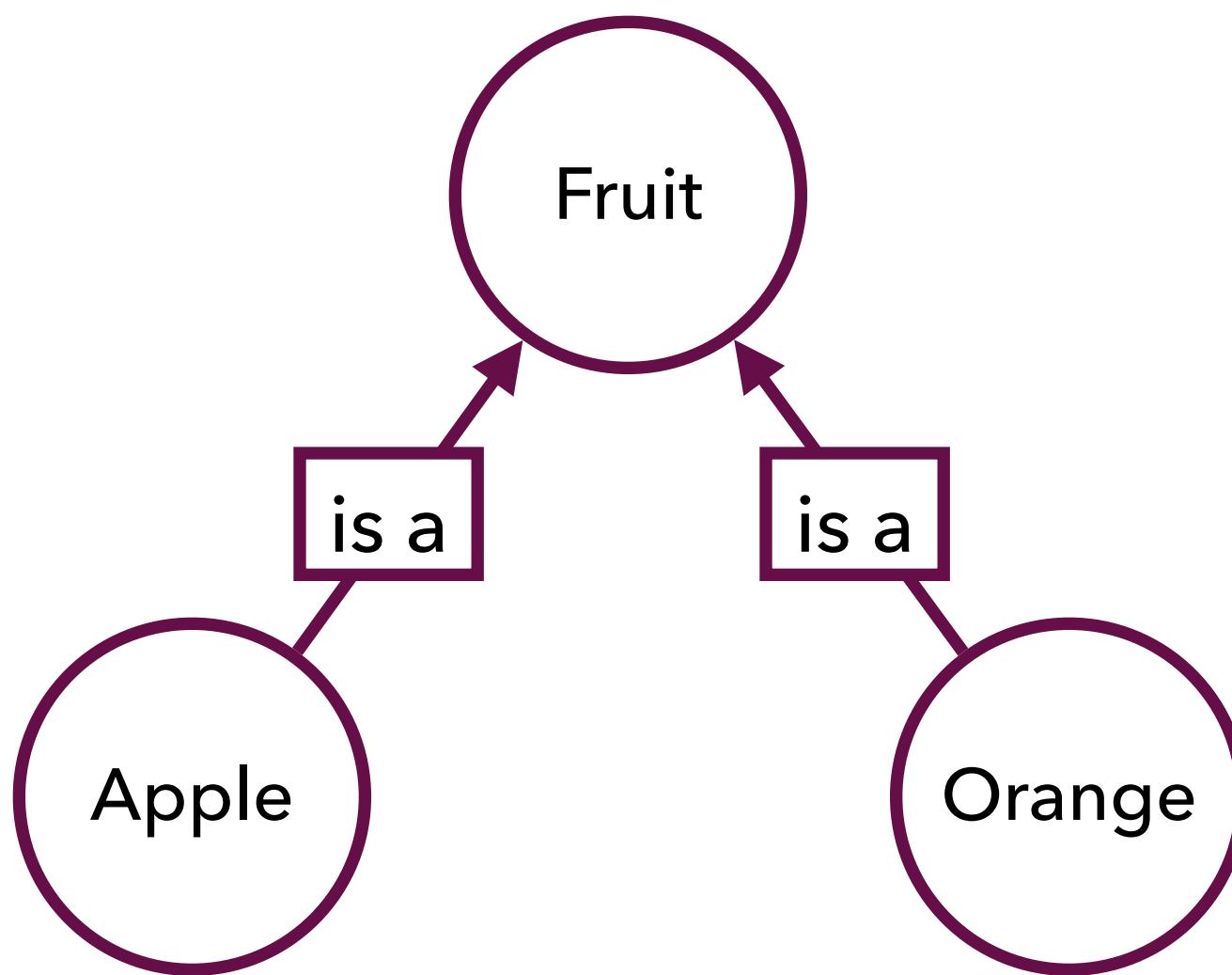
Interaction Graph

In many ways, graphs are the main modality of data we receive from nature.

- Petar Veličković

Why Graphs?

Symbolic Representation



Sub-symbolic Representation

Apple: [0.2345, 0.234, 0.85757, ...]
Orange: [0.1345, 0.674, 0.856567, ...]
Fruit: [0.2366, 0.234, 0.8534, ...]

youtube.com/watch?v=knDDGYHnnSI

In many ways, graphs are the main modality of data we receive from nature.

- Petar Veličković

Why Graphs?



<https://deepgraphlearning.github.io/project/wikidata5m>

An introduction to causal reasoning in health analytics (Zhang et al)

<https://towardsdatascience.com/visualizing-protein-networks-in-python-58a9b51be9d5>

Knowledge Graph

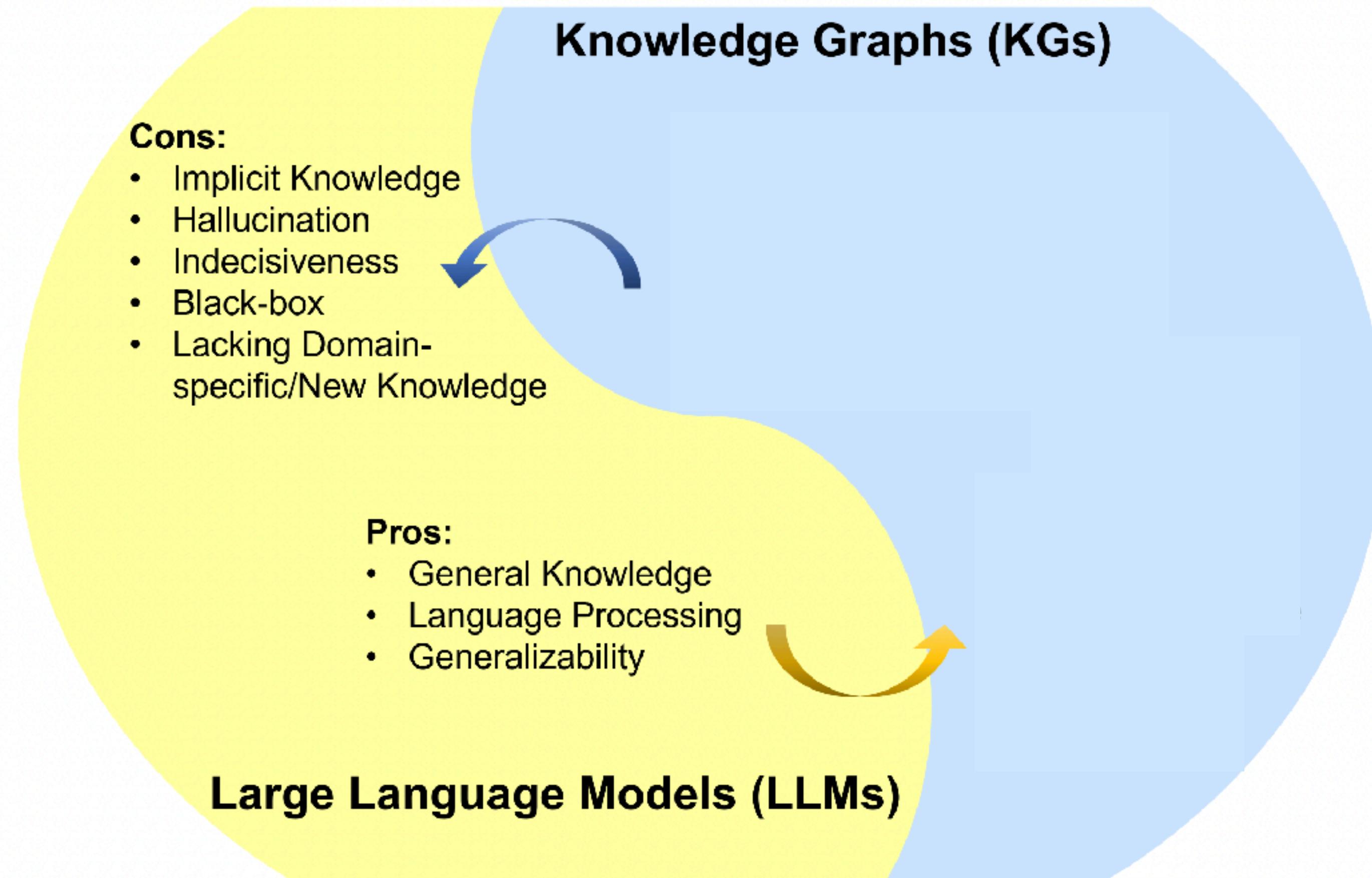
Causal Graph

Interaction Graph

In many ways, graphs are the main modality of data we receive from nature.

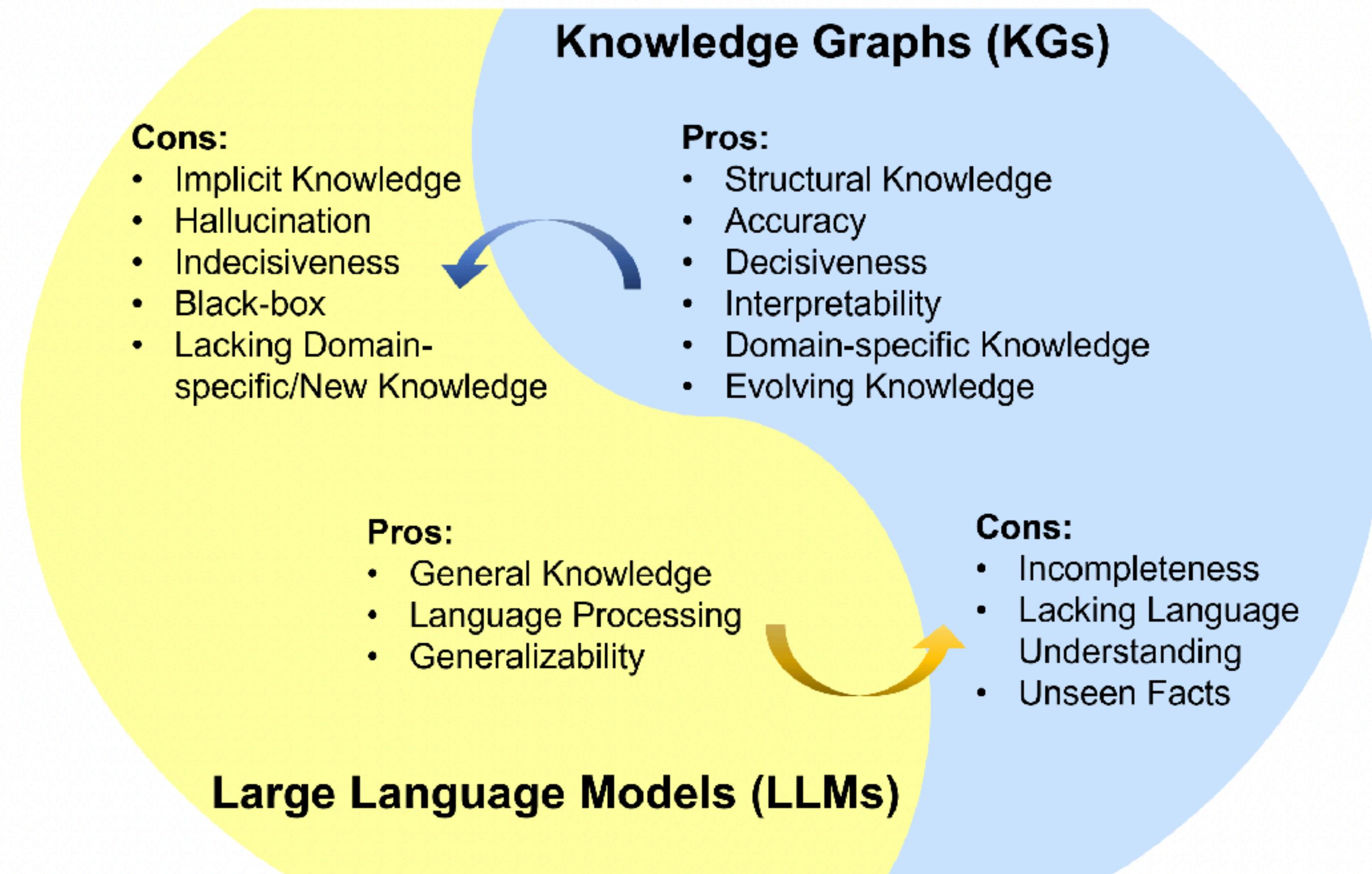
- Petar Veličković

Graphs and LLMs



Unifying Large Language Models and Knowledge Graphs: A Roadmap (Pan et al.)

Graphs and LLMs



Unifying Large Language Models and Knowledge Graphs: A Roadmap (Pan et al.)

Graphs and LLMs

Air Canada chatbot promised a discount. Now the airline has to pay it.

Air Canada argued the chatbot was a separate legal entity 'responsible for its own actions,' a Canadian tribunal said



By [Kyle Melnick](#)

February 18, 2024 at 8:35 p.m. EST

https://www.youtube.com/watch?v=DkbX8O9zd_8&ab_channel=ArangoDB

Thriving Research

1. arXiv:2408.16667 [pdf, other] cs.LG cs.NI cs.CL cs.MA

Iterative Graph Alignment

Authors: Fangyuan Yu, Hardeep Singh Arora, Matt Johnson

Abstract: By compressing diverse narratives, LLMs go beyond memorization, achieving intelligence by capturing generalizable causal relationships. However, they suffer from local 'representation gaps' due to insufficient training data diversity, limiting their real-world utility, especially in tasks requiring strict alignment to rules. Traditional alignment met...

Submitted 29 August, 2024; originally announced August 2024.

Comments: 12 pages, 4 figures

2. arXiv:2407.13989 [pdf, other] cs.LG cs.AI

Enhancing Data-Limited Graph Neural Networks by Actively Distilling Knowledge from Large Language Models

Authors: Quan Li, Tianxiang Zhao, Lingwei Chen, Junjie Xu, Suhang Wang

Abstract: Graphs are pervasive in the real-world, such as social network analysis, bioinformatics, and...

Submitted 28 August, 2024; v1 submitted 18 July, 2024; originally announced July 2024.

Comments: 10 pages, 3 figures

3. arXiv:2408.15903 [pdf, other] cs.CL

LLM-Based Multi-Hop Question Answering with Knowledge Graph Integration in Evolving Environments

Authors: Ruirui Chen, Weifeng Jiang, Chengwei Qin, Ishaan Singh Rawal, Cheston Tan, Dongkyu Choi, Bo Xiang, Bo Ai

Abstract: The rapid obsolescence of information in Large Language Models (LLMs) has driven the development of various techniques to incorporate new facts. However, existing methods for...

Submitted 28 August, 2024; originally announced August 2024.

4. arXiv:2405.07626 [pdf, other] cs.LG cs.AI

AnomalyLLM: Few-shot Anomaly Edge Detection for Dynamic Graphs using Large Language Models

Authors: Shuo Liu, Di Yao, Lanting Fang, Zhiqiao Li, Wenbin Li, Kaiyu Feng, Xiaowen Ji, Jingping Bi

Abstract: Detecting anomaly edges for dynamic graphs aims to identify edges significantly deviating from the normal pattern and can be applied in various domains, such as cybersecurity, financial transactions and AIOps. With the evolving of time, the types of anomaly edges are emerging and the labeled anomaly samples are few for each type. Current methods are either d...

Submitted 28 August, 2024; v1 submitted 13 May, 2024; originally announced May 2024.

Comments: 13 pages

5. arXiv:2408.07511 [pdf, other] cs.CL cs.R

WeKnow-RAG: An Adaptive Approach for Retrieval-Augmented Generation Integrating Web Search and Knowledge Graphs

Authors: Weijian Xie, Xuefeng Liang, Yuhui Liu, Kaihua Ni, Hong Cheng, Zetian Hu

Abstract: Large Language Models (LLMs) have greatly contributed to the development of adaptive intelligent agents and are positioned as an important way to achieve Artificial General Intelligence (AGI). However,...

Submitted 27 August, 2024; v1 submitted 14 August, 2024; originally announced August 2024.

Comments: 8 pages, 2 figures, technical report for 3rd place in Task 3 of Meta KDD Cup 2024 CRAG Challenge

6. arXiv:2408.14964 [pdf, other] cs.LG

Cross-Modal Learning for Chemistry Property Prediction: Large Language Models Meet Graph Machine Learning

Authors: Sakhinana Sagar Srinivas, Venkataramana Runkana

Abstract: ...is to create novel molecules with desired properties, facilitating accurate property predictions for applications such as material design and drug screening. However, existing graph deep learning methods face limitations that curb their expressive power. To address this, we explore the integration of vast molecular domain...

Submitted 27 August, 2024; originally announced August 2024.

Comments: Paper Accepted at Workshop on Deductive and Differentiable Reasoning in Generative Models at NeurIPS 2024

1. arXiv:2408.16667 [pdf, other] cs.LG cs.NI cs.CL cs.MA

Structured Data Analysis using Knowledge Graph + LLM

The post-2020 era is referred to as the digital era. Most companies are re-inventing their working model by digitizing...

Jan 20 · 259 · 3



2. arXiv:2407.13989 [pdf, other] cs.LG cs.AI

Ontotext Marketing Gets a Boost from Knowledge Graph Powered LLMs

An abbreviated and updated version of a presentation from Ontotext's Knowledge Graph Forum 2023 about the synergy...

Mar 13 · 14 · 1

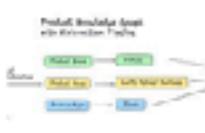


3. arXiv:2408.15903 [pdf, other] cs.CL

Build Product Knowledge Graph using LLM

Extracting attribute-value information from unstructured data is called named entity recognition. There are various ways of...

Aug 2 · 159 · 1



4. arXiv:2405.07626 [pdf, other] cs.LG cs.AI

Can you mention the tech stack to build knowledge graph based LLMs?

Nov 16, 2023 · 9 · 0



5. arXiv:2408.07511 [pdf, other] cs.CL cs.R

Navigating the World of Knowledge Graphs : Part 1

Knowledge Graph with LLM Story

Nov 20, 2023 · 116 · 1



6. arXiv:2408.14964 [pdf, other] cs.LG

It does now :) Take a look:

https://github.com/ahuinyk/knowledge_graph_maker/tree/main/knowledge_graph_maker/lm_clients

Jul 30



7. arXiv:2408.14964 [pdf, other] cs.LG

Large language models(LLM) & Knowledge Graph(KG) : 1+1>2 ?

Erik Protocol has always been committed to building robust



Thriving Research

1. arXiv:2408.16657 [pdf, other] cs.LG cs.AI cs.CL cs.MA
Iterative Graph Alignment
Authors: Fangyuan Yu, Hardeep Singh Arora, Matt Johnson
Abstract: By compressing diverse narratives, LLMs go beyond memorization, achieving intelligence by capturing generalizable causal relationships. However, they suffer from local 'representation gaps' due to insufficient training data diversity, limiting their real-world utility, especially in tasks requiring strict alignment to rules. Traditional alignment met... [More](#)
Submitted 29 August, 2024; originally announced August 2024.
Comments: 12 pages, 4 figures

2. arXiv:2407.13989 [pdf, other] cs.LG cs.AI
Enhancing Data-Limited Graph Neural Networks by Active Learning Knowledge from Large Language Models
Authors: Quan Li, Tianxiang Zhao, Lingwei Chen, Junjie Xu, Suhan Wang
Abstract: Graphs are pervasive in the real-world, such as social networks, biological networks, and... [More](#)
Submitted 28 August, 2024; v1 submitted 18 July, 2024; originally announced August 2024.
Comments: 10 pages, 3 figures

3. arXiv:2408.15903 [pdf, other] cs.CL
LLM-Based Multi-Hop Question Answering with Knowledge Graph Integration in Evolving Environments
Authors: Ruirui Chen, Weifeng Jiang, Chengwei Qin, Ishaan Singh Rawal, Cheston Tan, Dongkyu Choi, Bo Xiang, Bo Ai
Abstract: The rapid obsolescence of information in Large Language Models (LLMs) has driven the development of various techniques to incorporate new facts. However, existing methods for... [More](#)
Submitted 28 August, 2024, originally announced August 2024.

4. arXiv:2405.07626 [pdf, other] cs.LG cs.AI
AnomalyLLM: Few-shot Anomaly Edge Detection for Dynamic Graphs using Large Language Models
Authors: Shuo Liu, Di Yao, Lanting Fang, Zhetao Li, Wenbin Li, Ka yu Feng, XiaoWen Ji, Jingping Bi
Abstract: Detecting anomaly edges for dynamic graphs aims to identify edges significantly deviating from the normal pattern and can be applied in various domains, such as cybersecurity, financial transactions and AIOps. With the evolving of time, the types of anomaly edges are emerging and the labeled anomaly samples are few for each type. Current methods are either d... [More](#)
Submitted 28 August, 2024; v1 submitted 13 May, 2024; originally announced May 2024.
Comments: 13 pages

5. arXiv:2408.07611 [pdf, other] cs.CL cs.RL
WeKnow-RAG: An Adaptive Approach for Retrieval-Augmented Reinforcement Learning Integrating Web Search and Knowledge Graphs
Authors: Weijian Xie, Xuefeng Liang, Yuhui Liu, Kaihua Ni, Hongzhi Wang
Abstract: Large Language Models (LLMs) have greatly contributed to the development of adaptive intelligent agents and are positioned as one of the most promising tracks to achieve Artificial General Intelligence (AGI). However,... [More](#)
Submitted 27 August, 2024; v1 submitted 14 August, 2024; originally announced August 2024.
Comments: 8 pages, 2 figures, technical report for 3rd place in Task 3 of Meta KDD Cup 2024 CRAG Challenge

6. arXiv:2408.14954 [pdf, other] cs.LG
Cross-Modal Learning for Chemistry Property Prediction: Large Language Models Meet Graph Machine Learning
Authors: Sakhinana Sagar Srinivas, Venkataramana Runkana
Abstract: ...is to create novel molecules with desired properties, facilitating accurate property predictions for applications such as material design and drug screening. However, existing graph deep learning methods face limitations that curb their expressive power. To address this, we explore the integration of vast molecular domain... [More](#)
Submitted 27 August, 2024; originally announced August 2024.



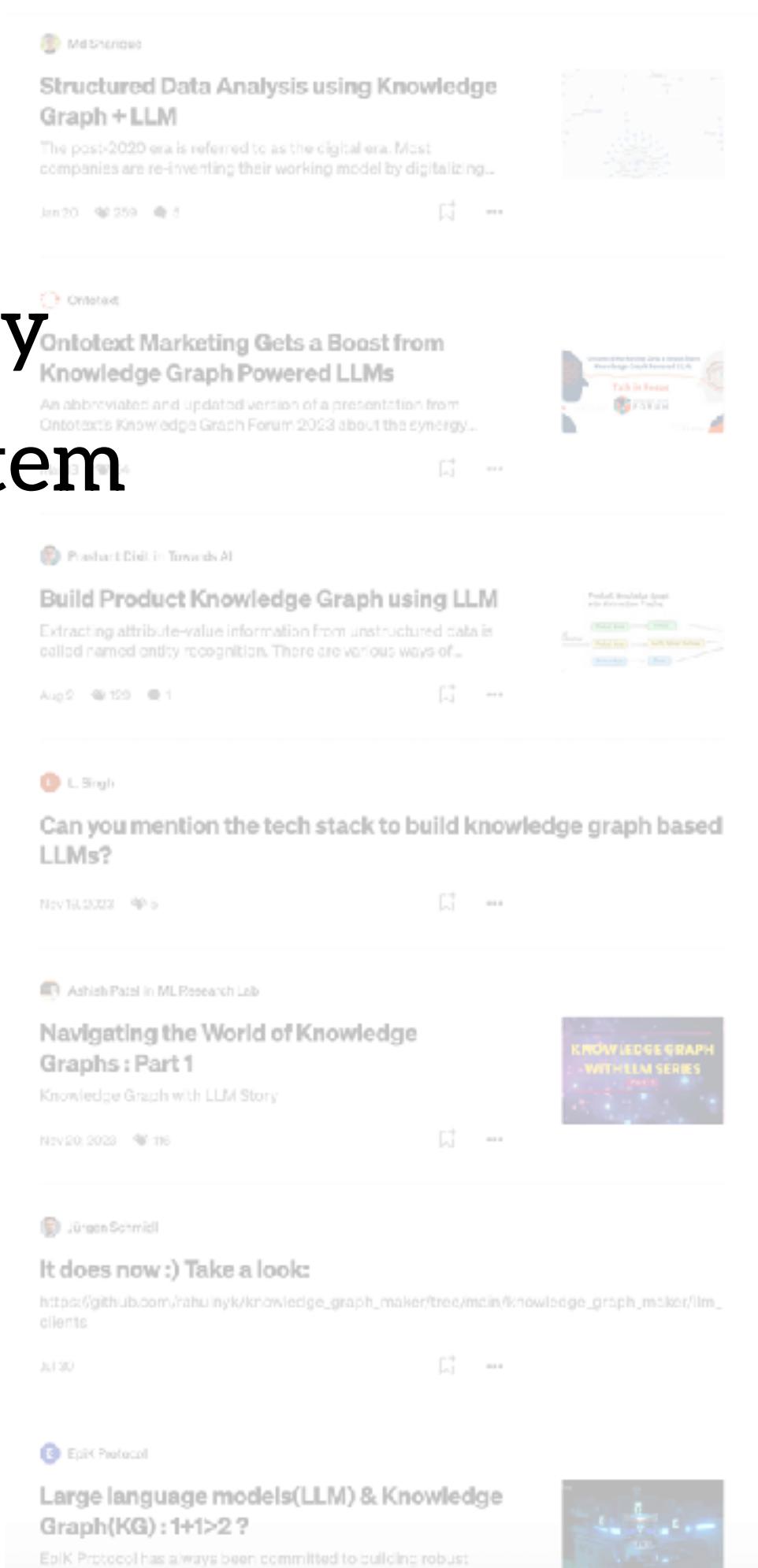
- New ideas every day
 - Functioning ecosystem
 - Easy to get started

Various techniques to incorporate new knowledge

 - Impossible to keep track of everything

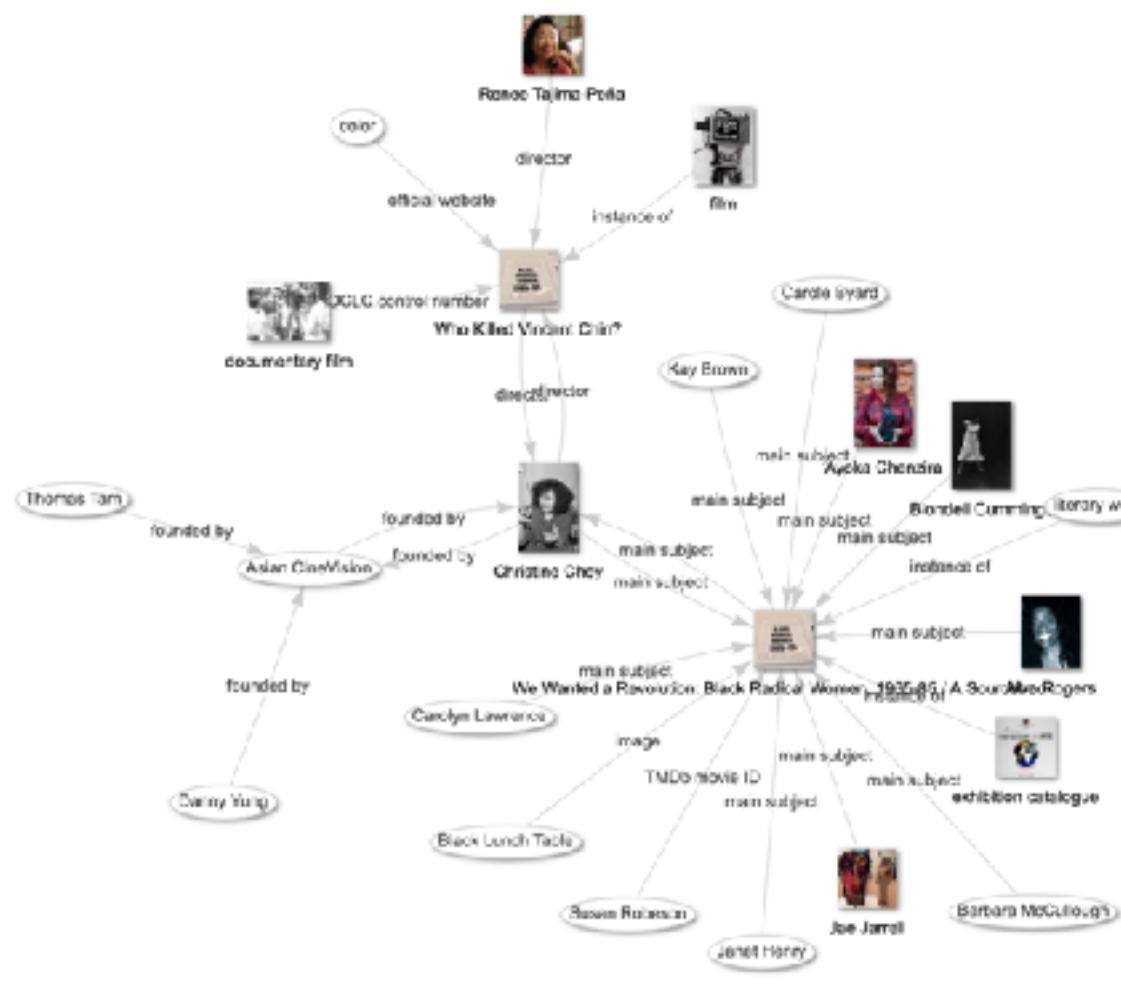


- Impossible to keep track of everything



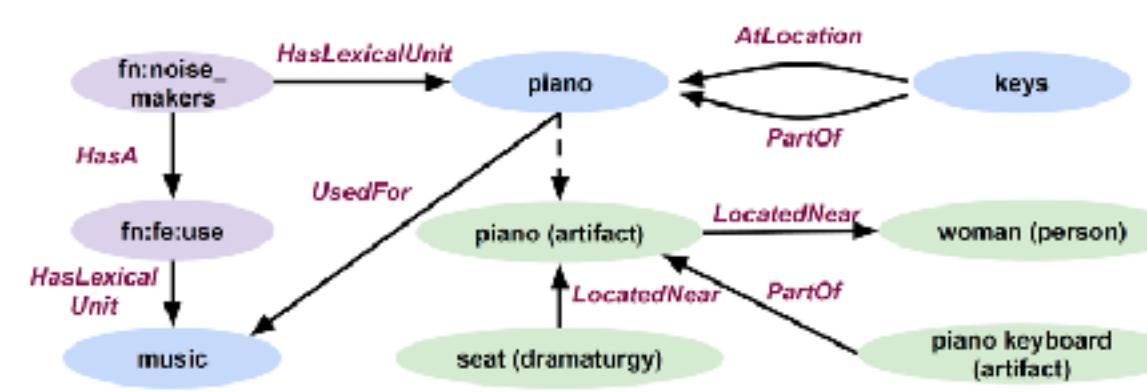
What is a Knowledge Graph?

A KG consists of entities (**nodes**) that are connected by relationships (**edges**).
Typically represented as list of (**subject, predicate, object**) triplets



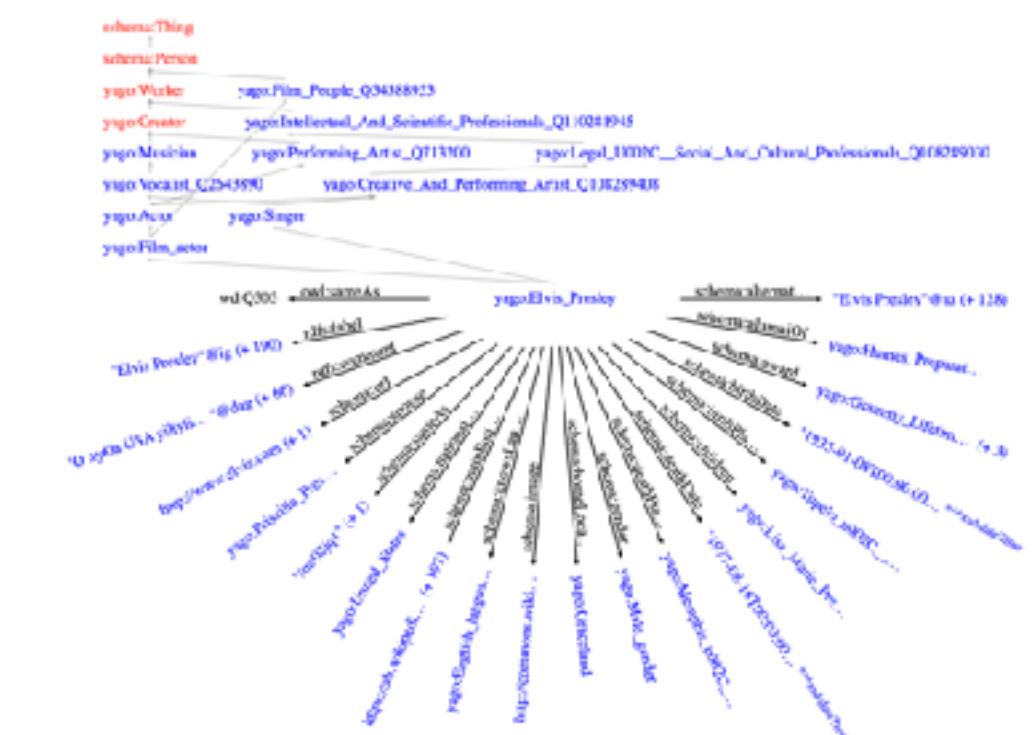
WikiData

> 100 million items



CSKG

> 2 million items



Yago

> 10 million items



DrugBank

20 thousand items

What is a Knowledge Graph?

A KG consists of entities (**nodes**) that are connected by relationships (**edges**).

Typically represented as list of (**subject, predicate, object**) triplets

Create private KG:



Financial Data



Medical data



Vacation Planning



Reading List

Taxonomy

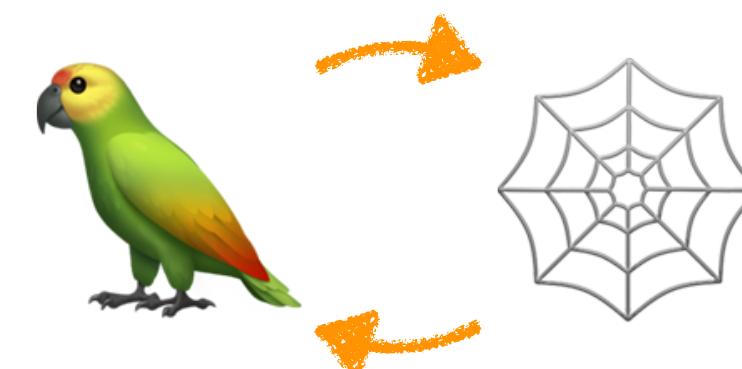
Goal



Improve LLM

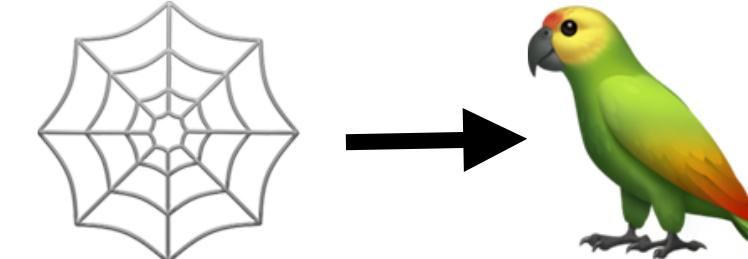


Improve KG

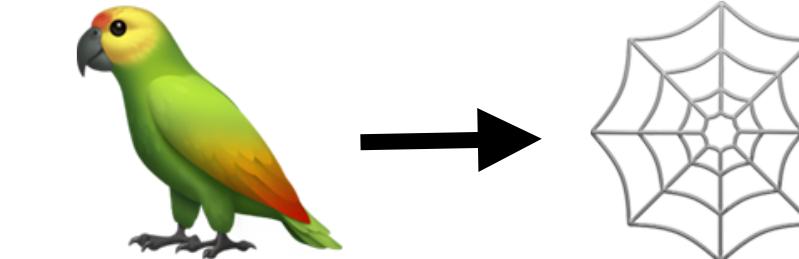


Synergy

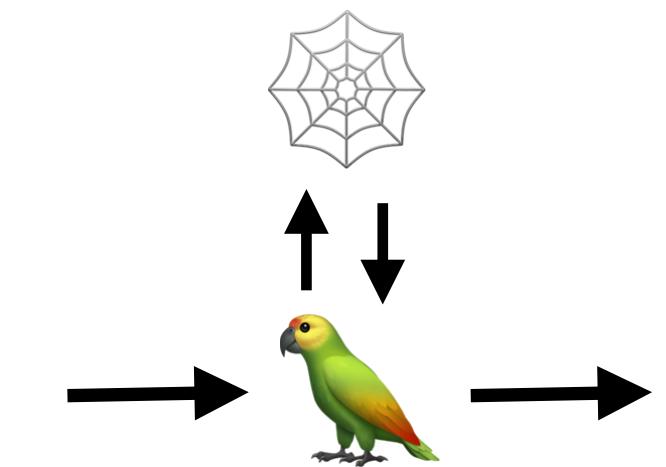
Input/Output



KG is Input



KG is Output

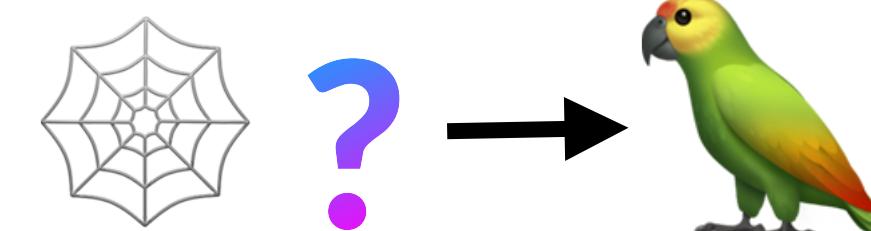


Reasoning using KG

Training/Inference



Use KG during Training



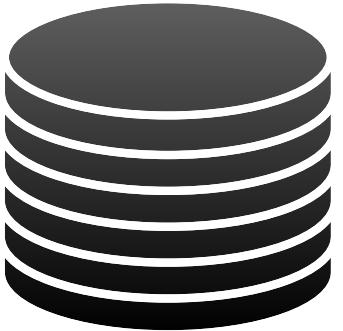
Use KG during Inference



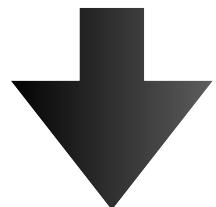
Part II

Some Great Ideas of KG and LLM Combinations

Constructing Graph Queries

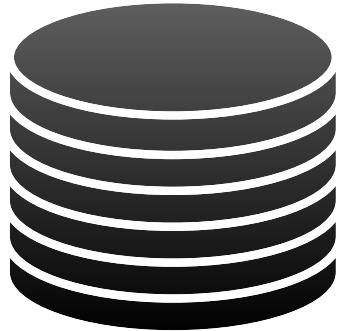


What are the names of the directors of all the movies Natalie Portman starred in between 2010 and 2020?

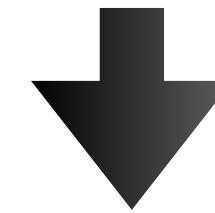


```
SELECT DISTINCT d.name AS director_name
FROM movies m
JOIN movie_cast mc ON m.movie_id = mc.movie_id
JOIN actors a ON mc.actor_id = a.actor_id
JOIN directors d ON m.director_id = d.director_id
WHERE a.name = 'Natalie Portman'
AND m.release_year BETWEEN 2010 AND 2020
AND mc.role = 'star';
```

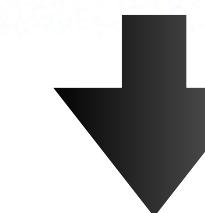
Constructing Graph Queries



What are the names of the directors of all the movies Natalie Portman starred in between 2010 and 2020?



```
SELECT DISTINCT d.name AS director_name
FROM movies m
JOIN movie_cast mc ON m.movie_id = mc.movie_id
JOIN actors a ON mc.actor_id = a.actor_id
JOIN directors d ON m.director_id = d.director_id
WHERE a.name = 'Natalie Portman'
    AND m.release_year BETWEEN 2010 AND 2020
    AND mc.role = 'star';
```

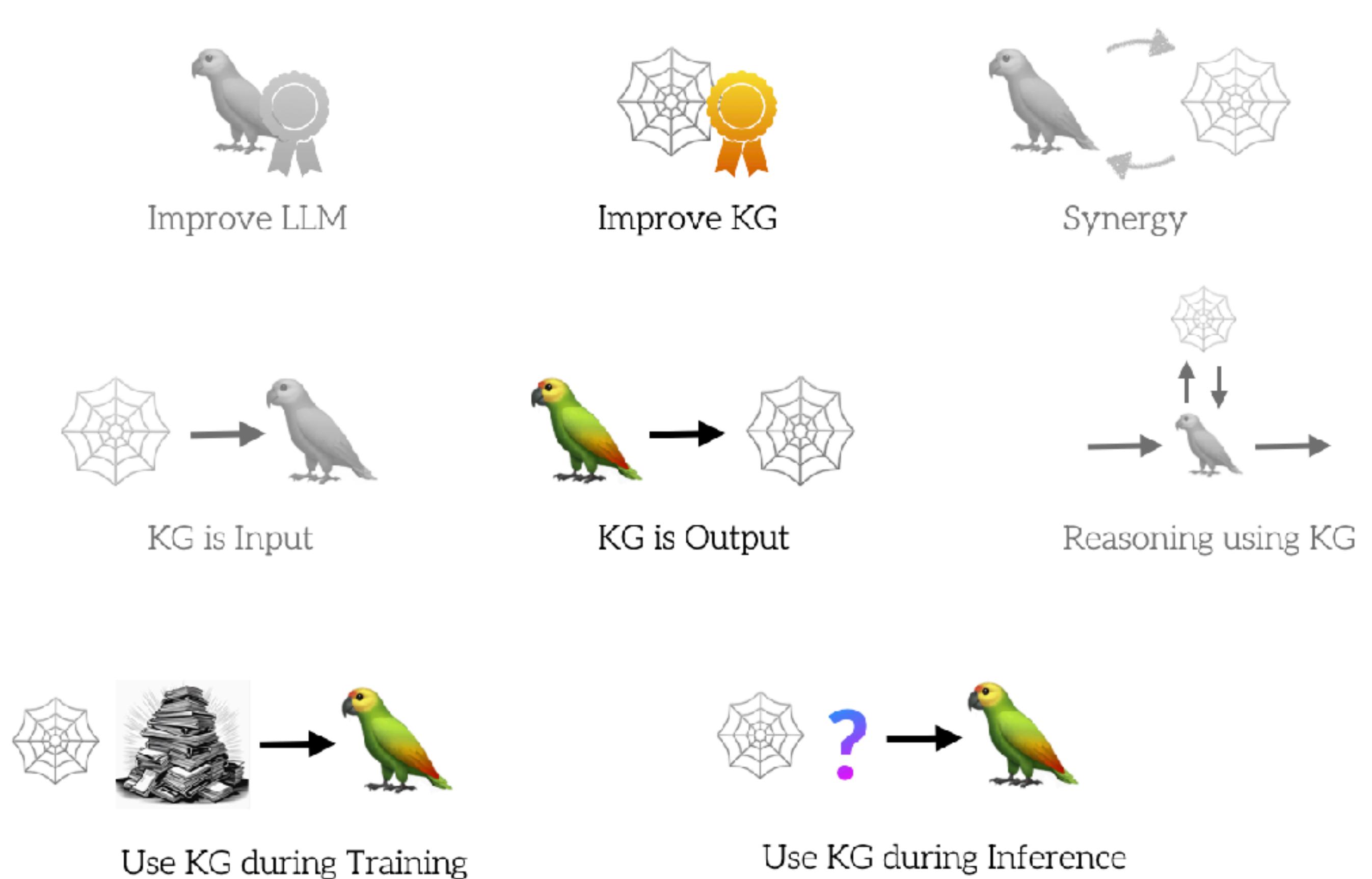


```
MATCH (a:Actor {name: "Natalie Portman"})-[:ACTED_IN]->
(m:Movie)-[:DIRECTED_BY]->(d:Director)
WHERE m.release_year >= 2010 AND m.release_year <= 2020
RETURN DISTINCT d.name AS director_name;
```

KG Completion

Exploring Large Language Models for Knowledge Graph Completion (Yao et al., 2024)

- KGs are (usually) incomplete
- Use LLMs to identify and generate missing links



KG Completion

Exploring Large Language Models for Knowledge Graph Completion (Yao et al., 2024)

Triplet classification: <Steve Jobs, founded, Apple Inc.> => **correct** or **incorrect**

Prompt: *Is this true: Steve Jobs founded Apple Inc.?*

KG Completion

Exploring Large Language Models for Knowledge Graph Completion (Yao et al., 2024)

Triplet classification: <Steve Jobs, founded, Apple Inc.> => **correct** or **incorrect**

Prompt: *Is this true: Steve Jobs founded Apple Inc.?*

Entity prediction: <Steve Jobs, ?, Apple Inc.> => **founded**

Prompt: *What is the relationship between Steve Jobs and Apple Inc.?*

Please choose your answer from:

was born in / founded / is citizen of / / plays for

KG Completion

Exploring Large Language Models for Knowledge Graph Completion (Yao et al., 2024)

Triplet classification: <Steve Jobs, founded, Apple Inc.> => **correct** or **incorrect**

Prompt: *Is this true: Steve Jobs founded Apple Inc.?*

Entity prediction: <Steve Jobs, ?, Apple Inc.> => **founded**

Prompt: *What is the relationship between Steve Jobs and Apple Inc.?*

Please choose your answer from:

was born in / founded / is citizen of / / plays for

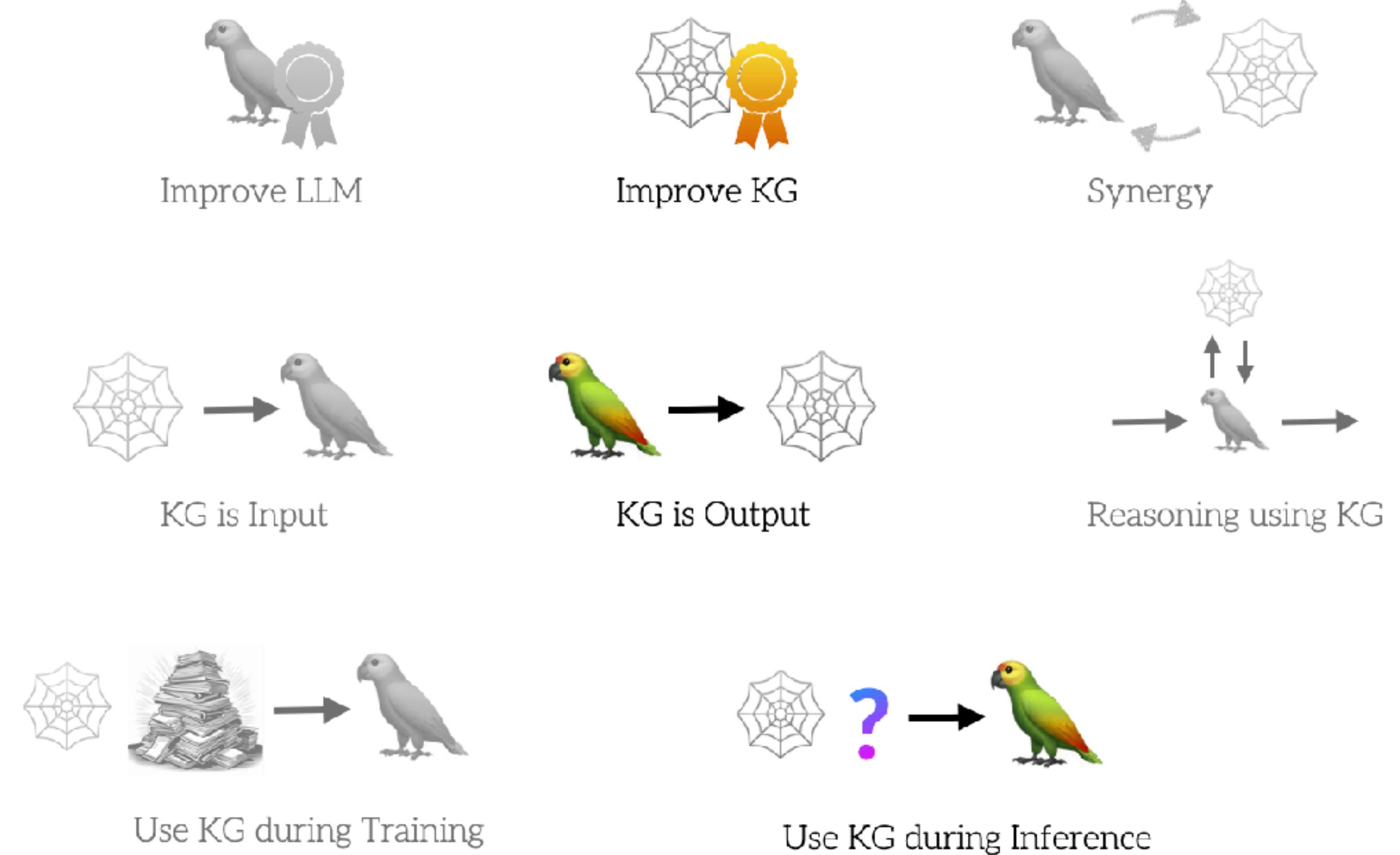
Fine-tuning/Instruction tuning
significantly improves results:

<Steve Jobs, founded, Apple Inc.> → **This is correct.**

KG Construction

Iterative Zero-Shot LLM Prompting for Knowledge Graph Construction (Carta et al., 2023)

- Creating comprehensive and accurate KGs is costly.
- Use zero-shot LLM prompting to extract and organize knowledge.



KG Construction

Iterative Zero-Shot LLM Prompting for Knowledge Graph Construction (Carta et al., 2023)

Candidate Triplet Extraction:

Identify relevant entities and extract their descriptions and type from the text.

(Greatly simplified)

KG Construction

Iterative Zero-Shot LLM Prompting for Knowledge Graph Construction (Carta et al., 2023)

Candidate Triplet Extraction:

Identify relevant entities and extract their descriptions and type from the text.

Resolve ambiguities:

Group the following entities into clusters based on their semantic similarity: Paris, Italy, car, motorbike

(Greatly simplified)

KG Construction

Iterative Zero-Shot LLM Prompting for Knowledge Graph Construction (Carta et al., 2023)

Candidate Triplet Extraction:

Identify relevant entities and extract their descriptions and type from the text.

Resolve ambiguities:

Group the following entities into clusters based on their semantic similarity: Paris, Italy, car, motorbike

Identify schema:

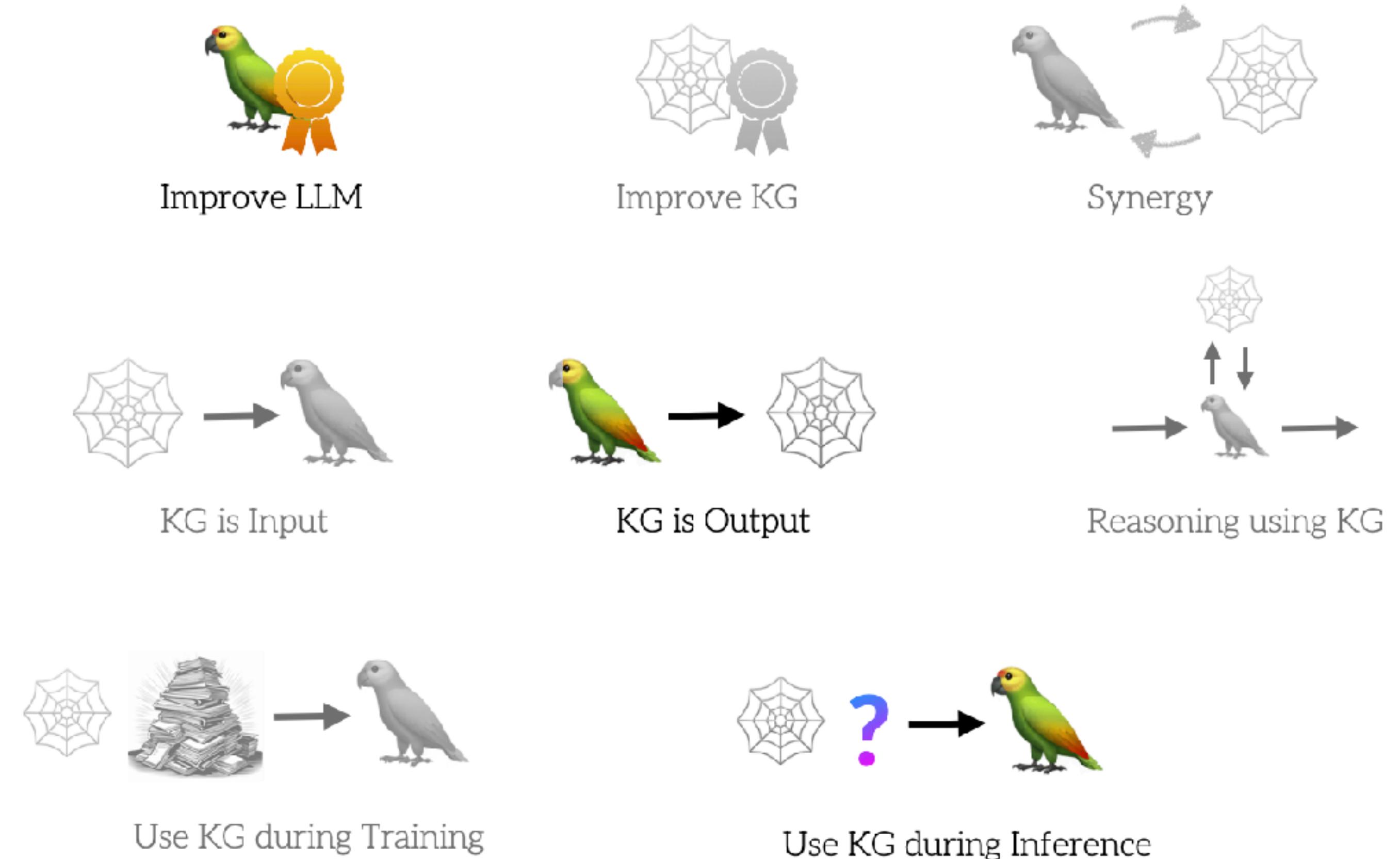
For the following set of entity types, generate a hypernym that can serve as a common category for the entire group: fish, crustacean

(Greatly simplified)

LLM Evaluation

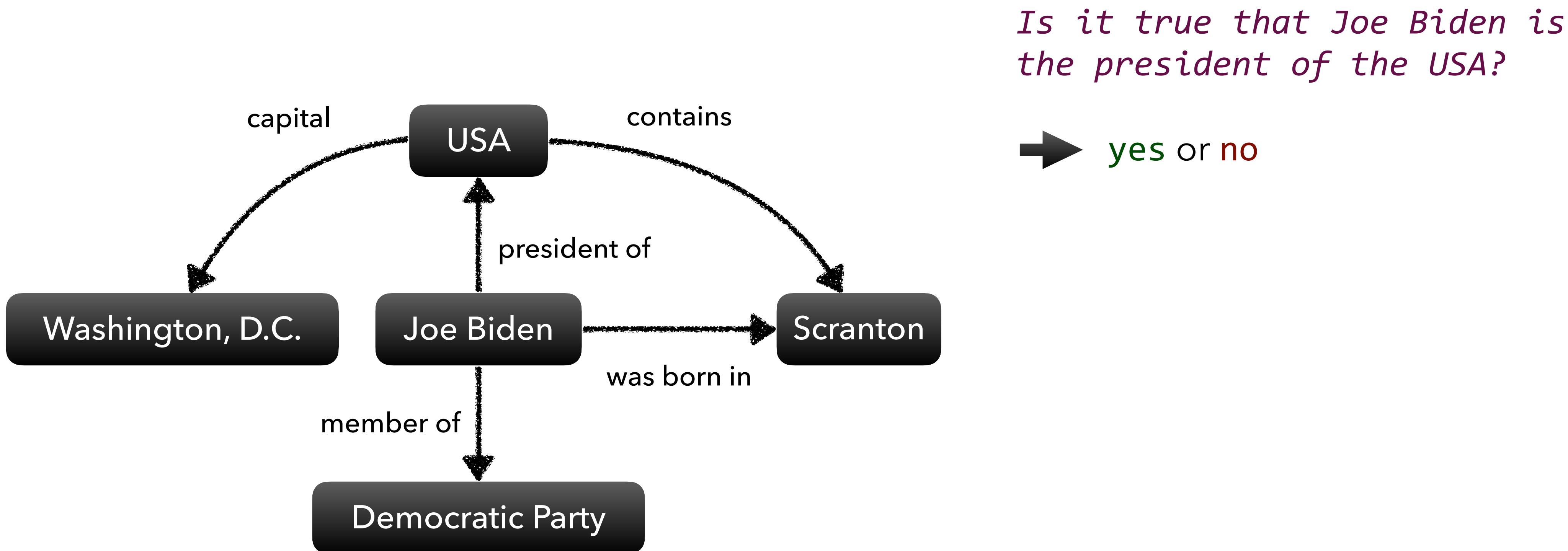
KGLens: Towards Efficient and Effective Knowledge Probing of Large Language Models with Knowledge Graphs (Zheng et al., 2023)

- Evaluating the factualness of LLMs is difficult to scale and adopt.
- Translate KG into natural language questions.
- Identify knowledge blind spots.



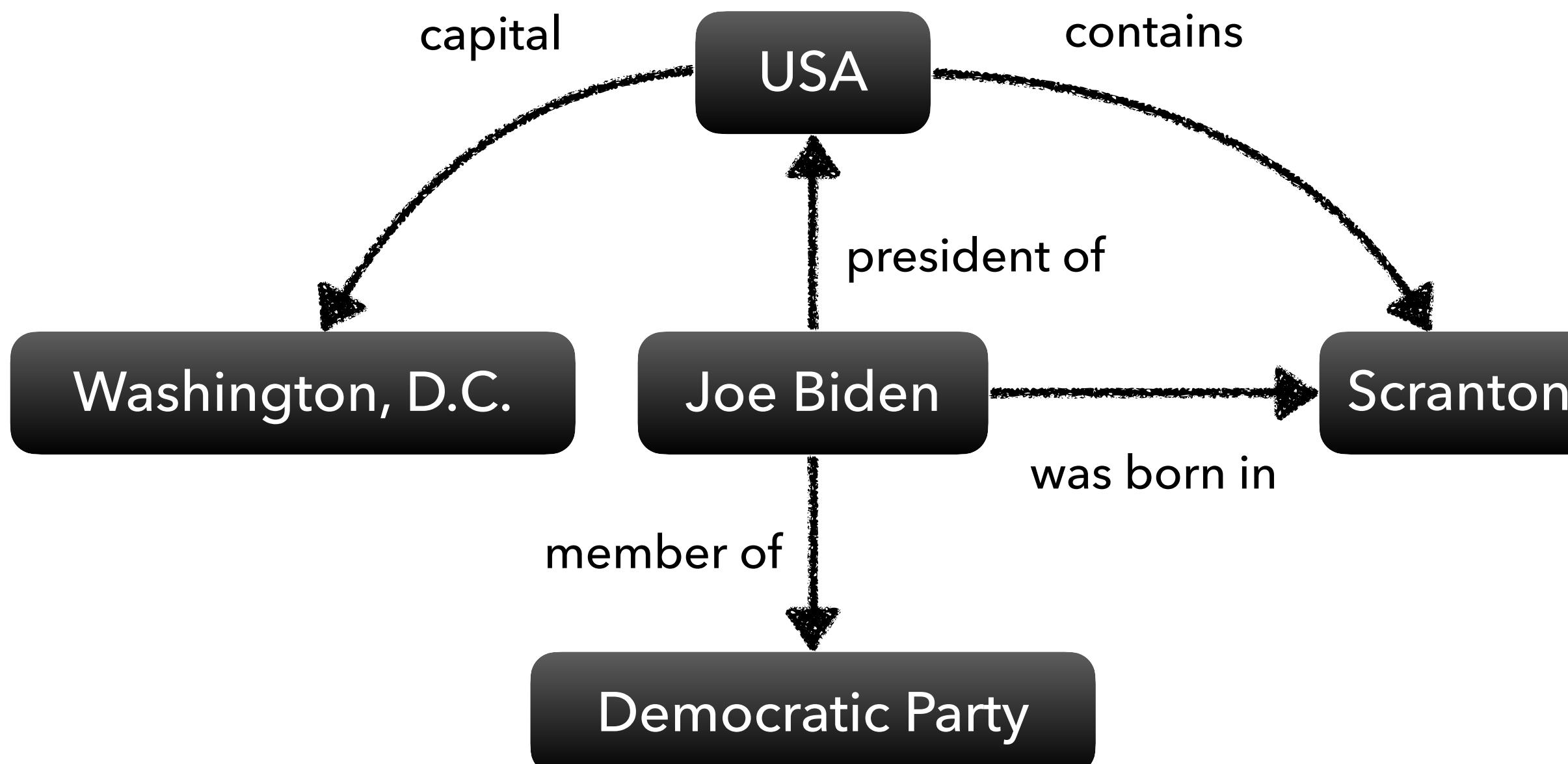
LLM Evaluation

KGLens: Towards Efficient and Effective Knowledge Probing of Large Language Models with Knowledge Graphs (Zheng et al., 2023)



LLM Evaluation

KGLens: Towards Efficient and Effective Knowledge Probing of Large Language Models with Knowledge Graphs (Zheng et al., 2023)



Is it true that Joe Biden is the president of the USA?

→ yes or no

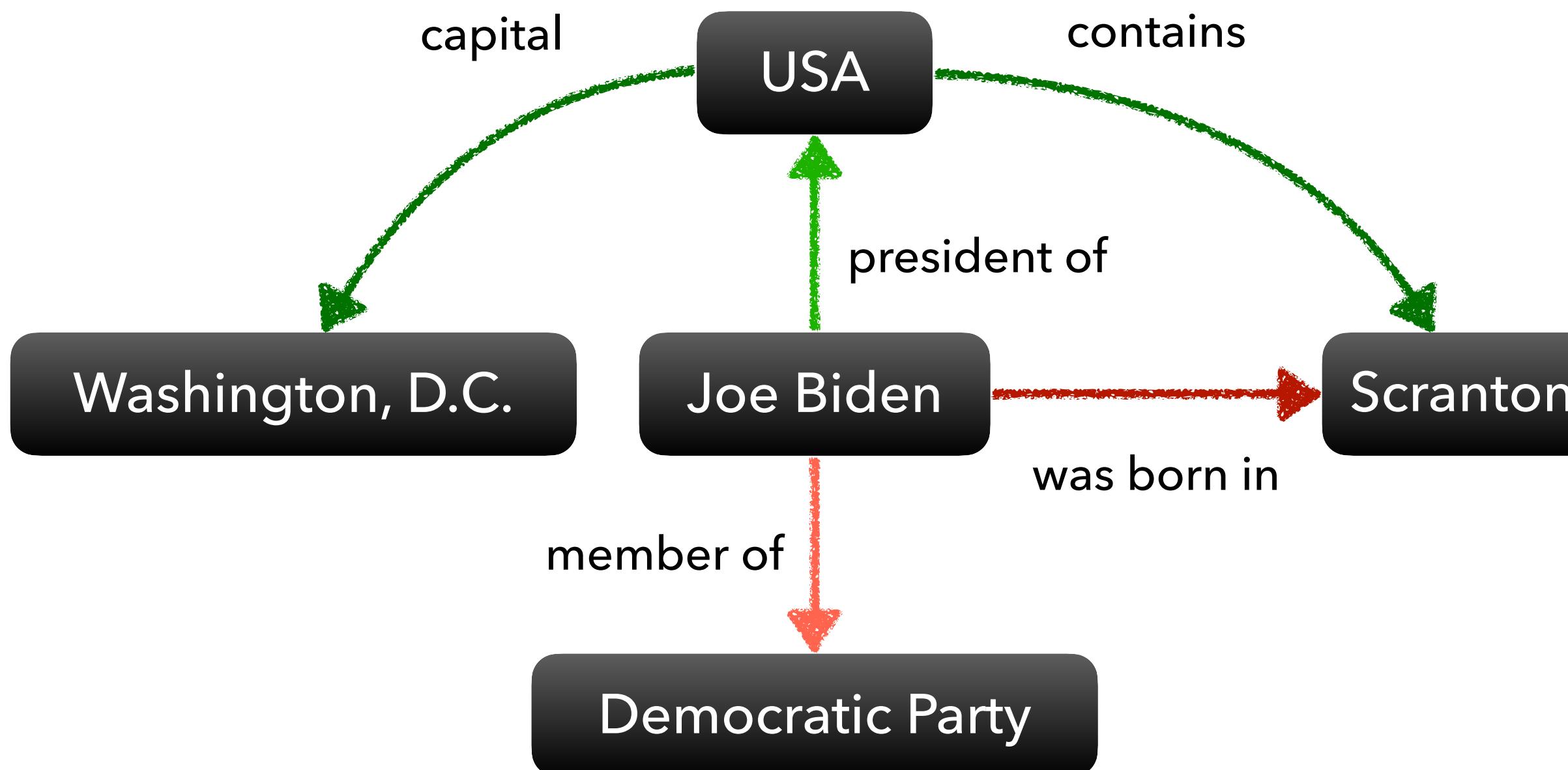
What is the capital of the USA?

→ The capital of the United States of America is Washington, D.C., which stands for the District of Columbia.

- Use LLM to check if the answer matches KG.

LLM Evaluation

KGLens: Towards Efficient and Effective Knowledge Probing of Large Language Models with Knowledge Graphs (Zheng et al., 2023)

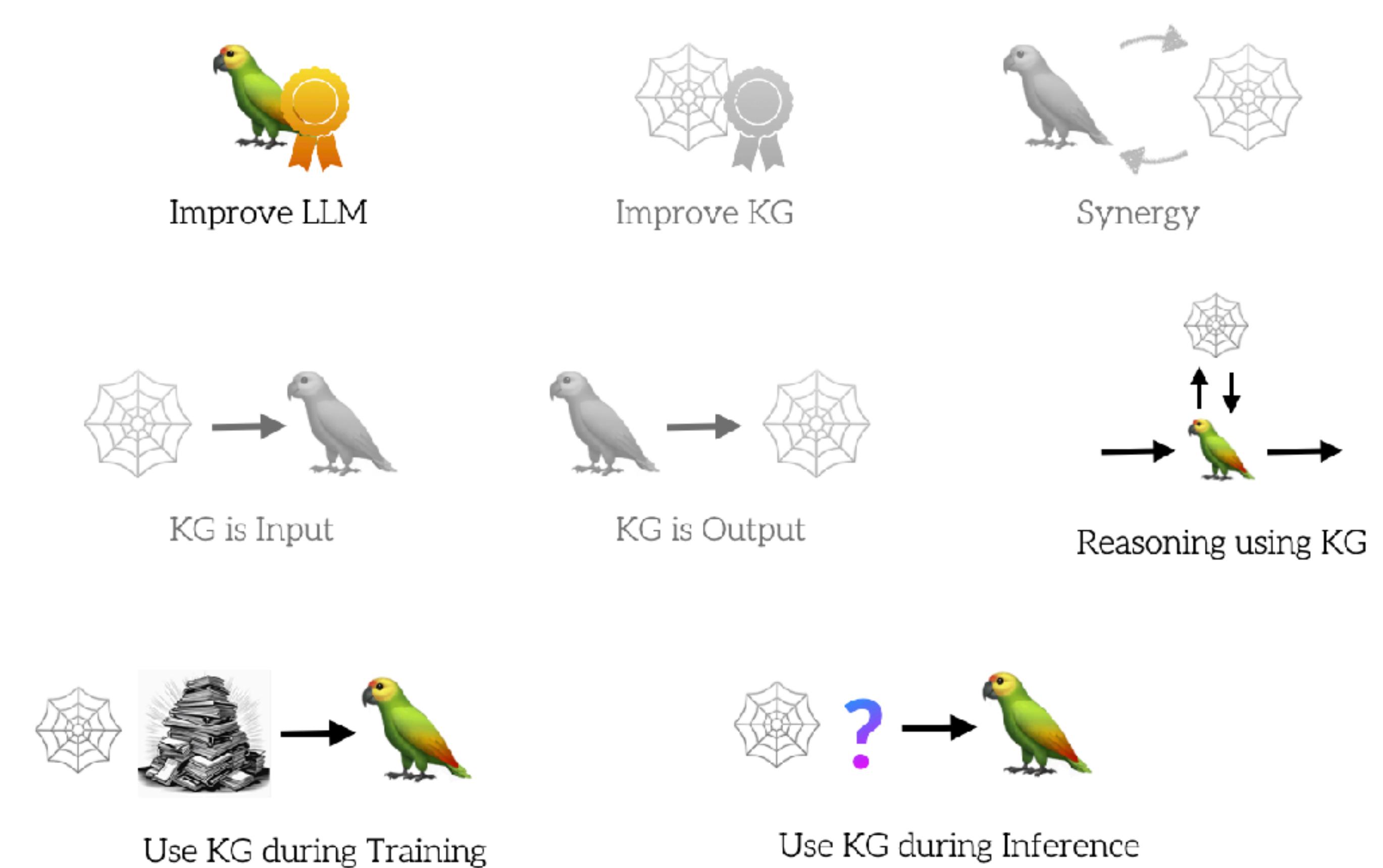


- Estimate likelihood of an LLM failing on each edge.
- Prioritize difficult edges.

Reasoning on Graphs

Reasoning on Graphs: Faithful and Interpretable Large Language Model Reasoning (Luo et al., 2024)

- Lack of (up-to-date) knowledge and hallucinations
- Ground LLM reasoning using KGs



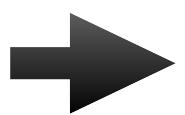
Reasoning on Graphs

Reasoning on Graphs: Faithful and Interpretable Large Language Model Reasoning (Luo et al., 2024)

What is the nationality of Joe Biden?

Generate plan:

*Please generate helpful
relation paths for answering
the question: "What is the
nationality of Joe Biden?"*



<Name> → born in → ? → city of → ?

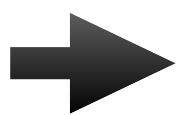
Reasoning on Graphs

Reasoning on Graphs: Faithful and Interpretable Large Language Model Reasoning (Luo et al., 2024)

What is the nationality of Joe Biden?

Generate plan:

*Please generate helpful
relation paths for answering
the question: "What is the
nationality of Joe Biden?"*



<Name> → born in → ? → city of → ?

Retrieve reasoning paths from KGs:

Joe Biden → born in → **Scranton** → city of → **USA**

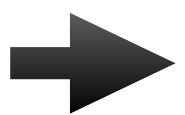
Reasoning on Graphs

Reasoning on Graphs: Faithful and Interpretable Large Language Model Reasoning (Luo et al., 2024)

What is the nationality of Joe Biden?

Generate plan:

*Please generate helpful
relation paths for answering
the question: "What is the
nationality of Joe Biden?"*



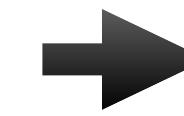
<Name> → born in → ? → city of → ?

Retrieve reasoning paths from KGs:

Joe Biden → born in → **Scranton** → city of → **USA**

Answer based on paths:

*Please answer the
question based on the
reasoning paths and
explain why.*



**Joe Biden's nationality
is American since he was
born in the USA.**

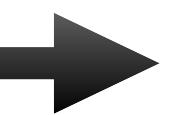
Reasoning on Graphs

Reasoning on Graphs: Faithful and Interpretable Large Language Model Reasoning (Luo et al., 2024)

What is the nationality of Joe Biden?

Generate plan:

*Please generate helpful
relation paths for answering
the question: "What is the
nationality of Joe Biden?"*

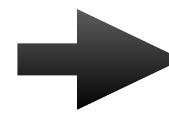


<Name> → born in → ? → city of → ?

Use fine-tuning to improve results.

Answer based on paths:

*Please answer the
question based on the
reasoning paths and
explain why.*

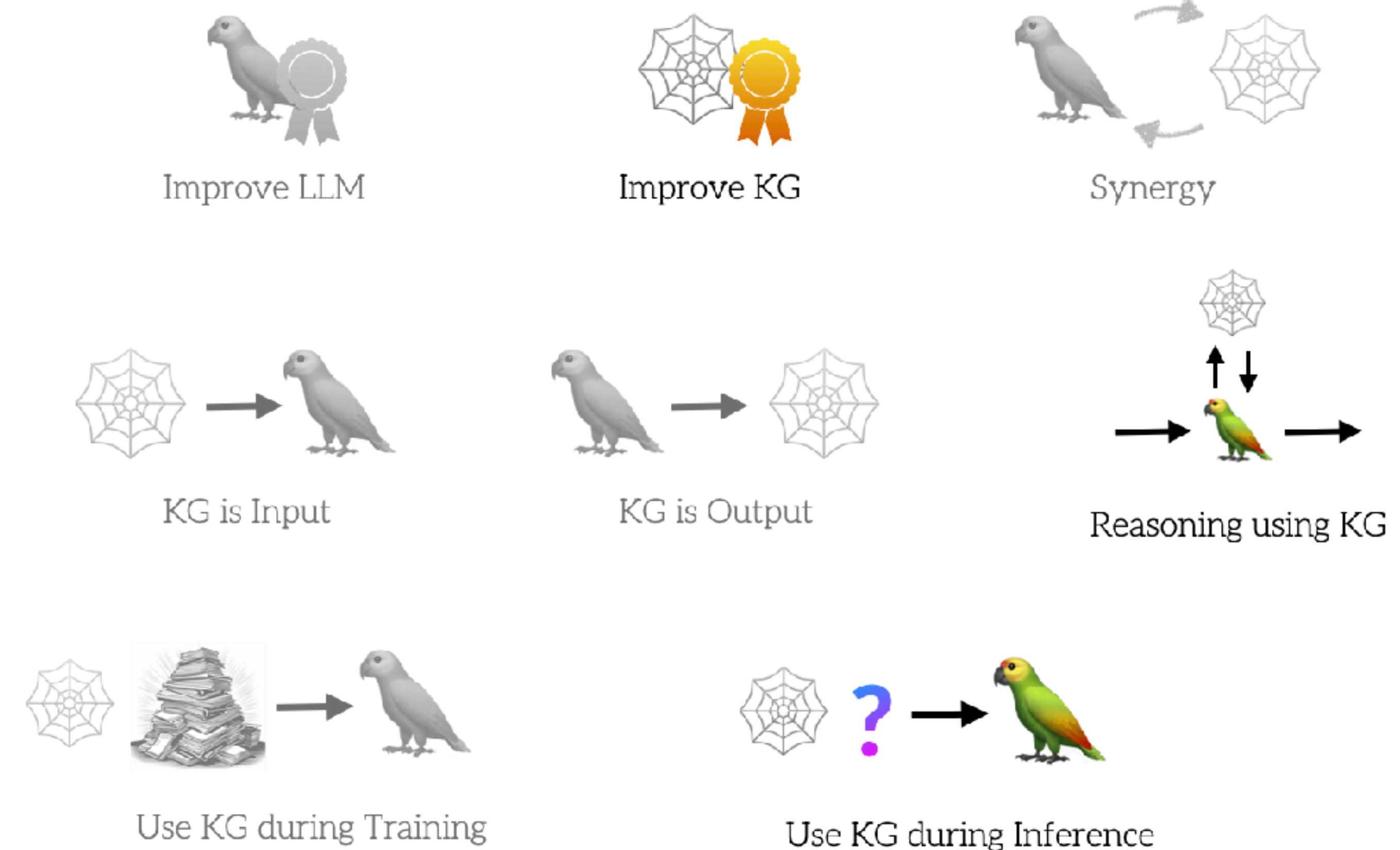


**Joe Biden's nationality
is American since he was
born in the USA.**

Reasoning on Graphs II

Graph Language Models (Plenz et al., 2024)

- Feeding KGs as triplets to LLMs is very inefficient
- Combine Graph Neural Network with LLM

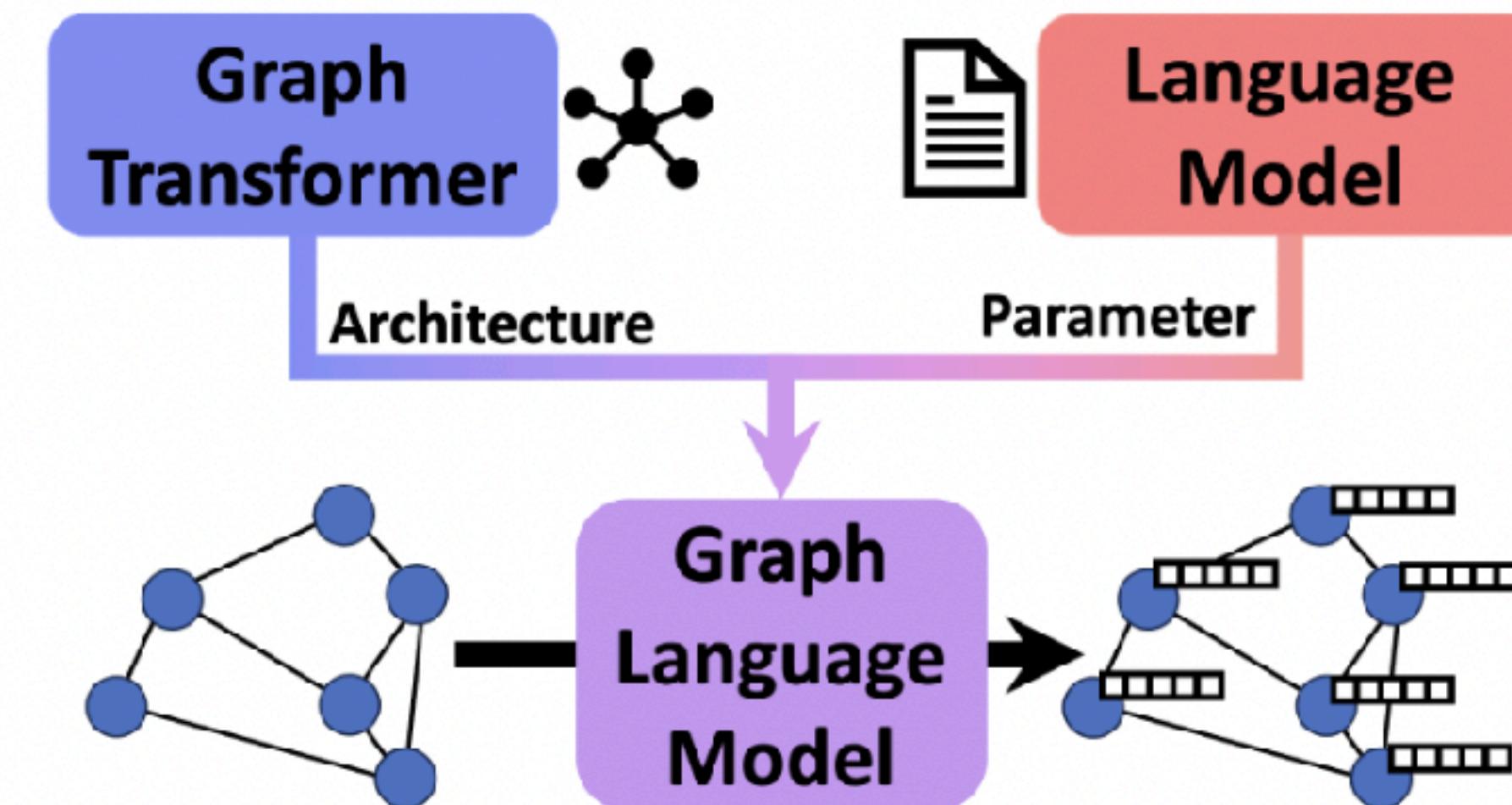


Reasoning on Graphs II

Graph Language Models (Plenz et al., 2024)



"[W]e transform the LM to a Graph Transformer (GT) – while maintaining compatibility with its pretrained LM parameters."

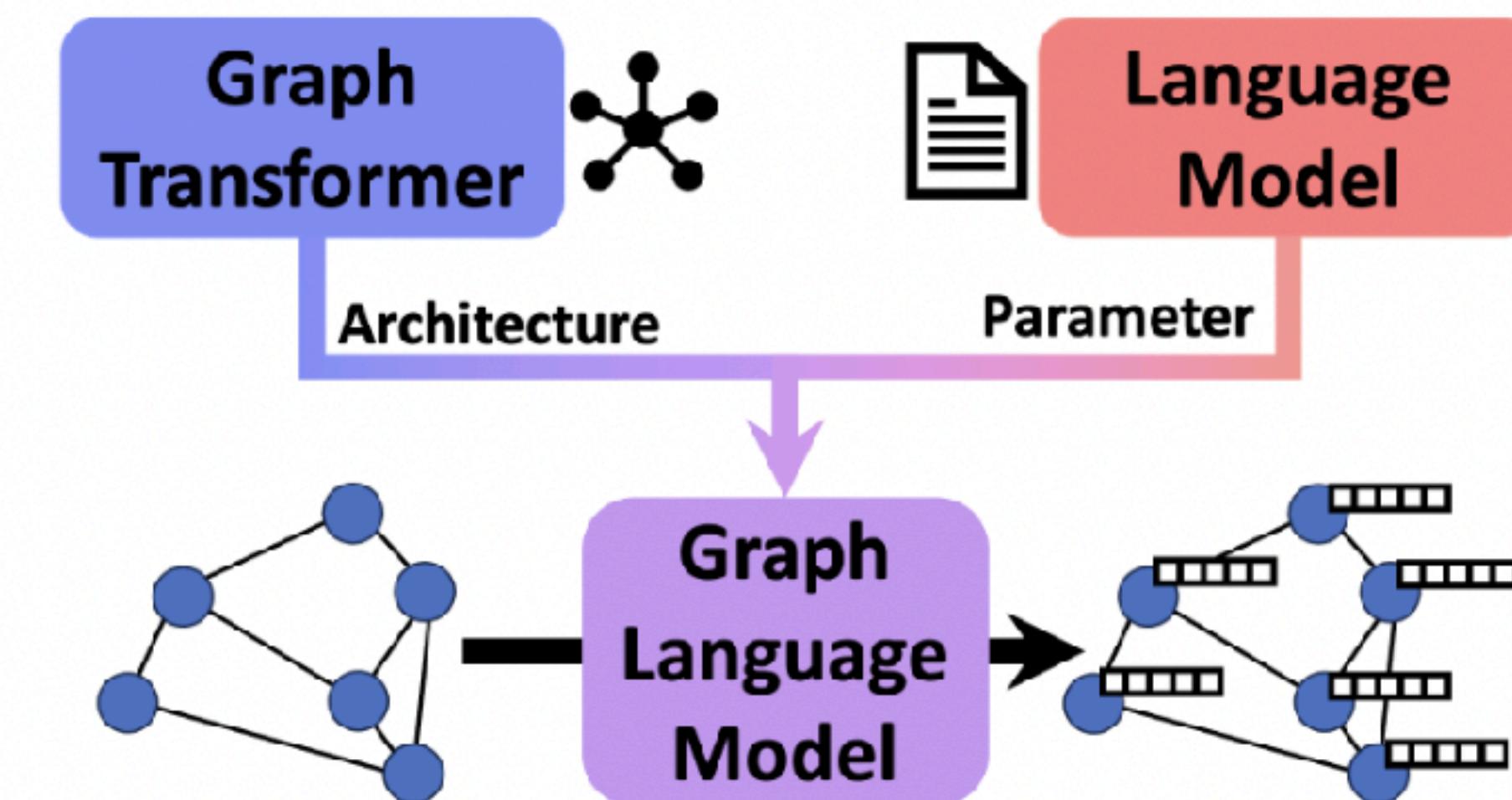


Reasoning on Graphs II

Graph Language Models (Plenz et al., 2024)

“[W]e transform the LM to a Graph Transformer (GT) – while maintaining compatibility with its pretrained LM parameters.”

- Add a GNN mechanism to a LLM
 - ▶ requires graph-based positional embeddings
- Use GNN + LLM parameters to reason over KG



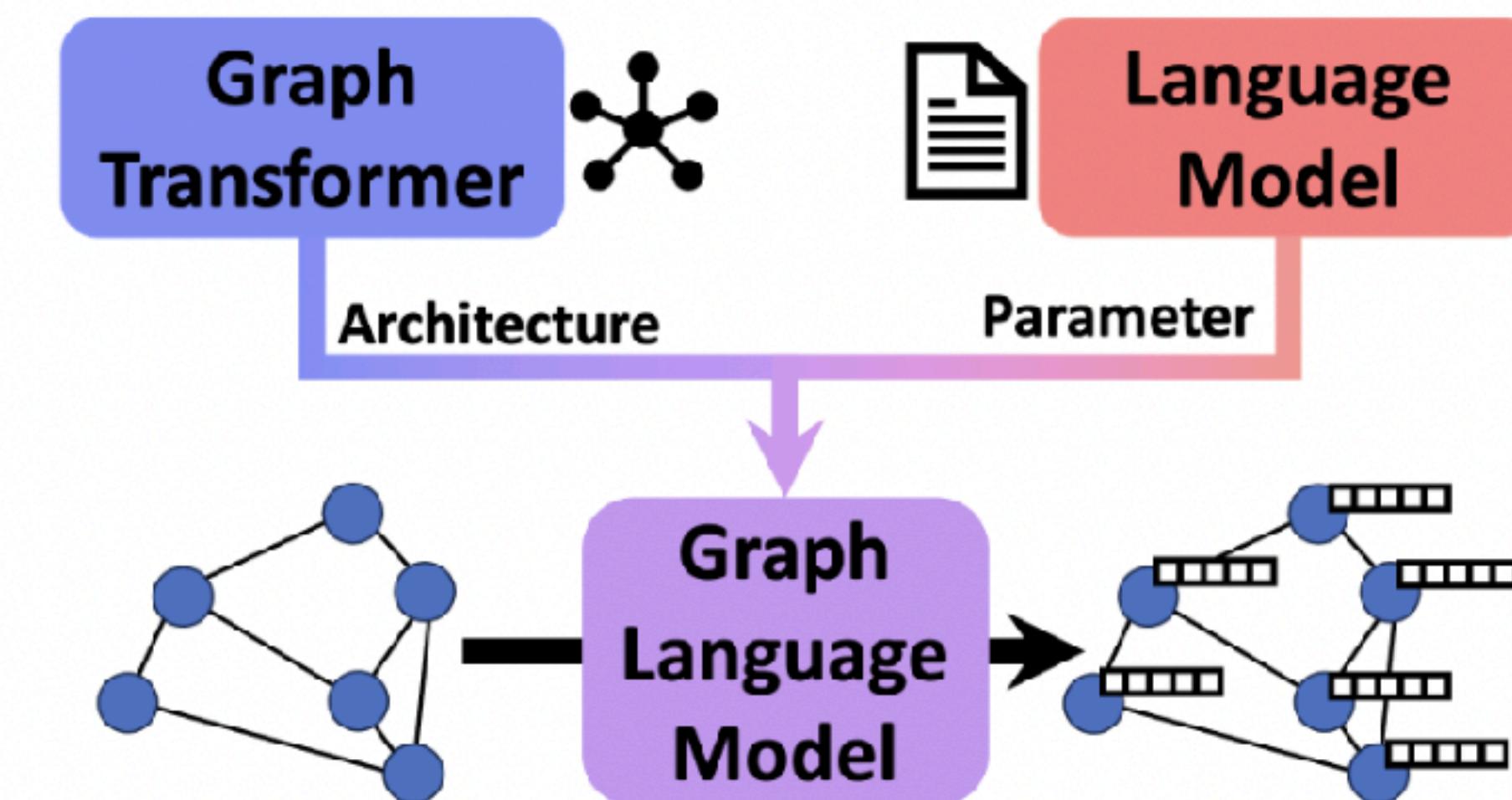
Entity prediction: <Steve Jobs, ?, Apple Inc.> => founded

Reasoning on Graphs II

Graph Language Models (Plenz et al., 2024)

“[W]e transform the LM to a Graph Transformer (GT) – while maintaining compatibility with its pretrained LM parameters.”

- Add a GNN mechanism to a LLM
 - ▶ requires graph-based positional embeddings
- Use GNN + LLM parameters to reason over KG
- T5 (only use the encoder)

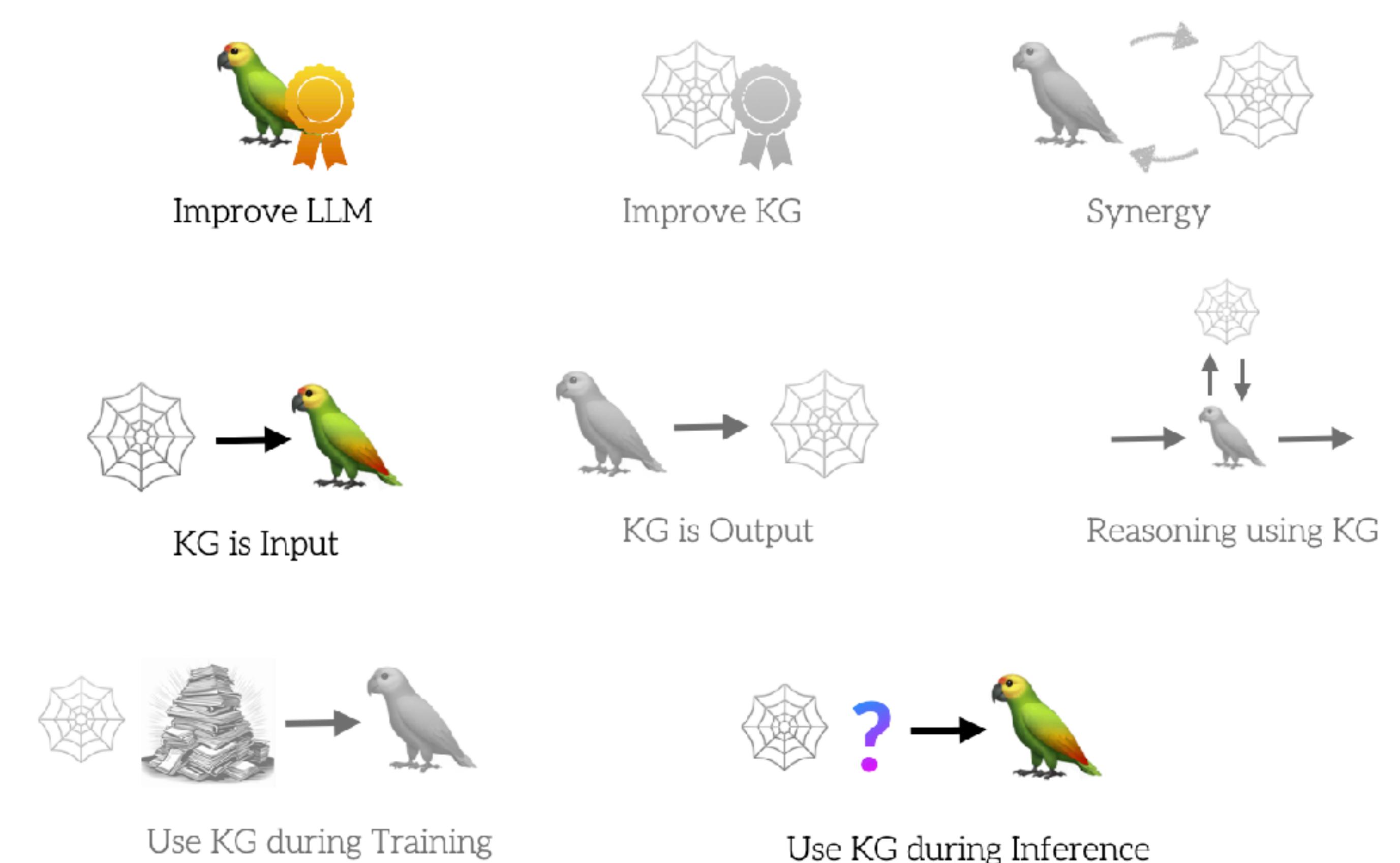


Entity prediction: <Steve Jobs, ?, Apple Inc.> => founded

Retrofitting KGs

Mitigating Large Language Model Hallucinations via Autonomous Knowledge Graph-based Retrofitting
(Guan et al., 2023)

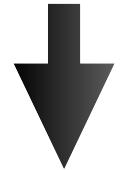
- Verify the LLM response based on a KG.
- Adjust it if incorrect.



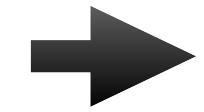
Retrofitting KGs

Mitigating Large Language Model Hallucinations via Autonomous Knowledge Graph-based Retrofitting
(Guan et al., 2023)

What is the capital of the USA?



The capital of the USA is NYC.

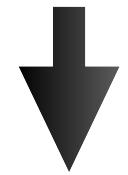


New York City → is capital of → **USA**

Retrofitting KGs

Mitigating Large Language Model Hallucinations via Autonomous Knowledge Graph-based Retrofitting
(Guan et al., 2023)

What is the capital of the USA?



The capital of the USA is NYC.

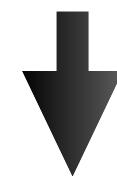


New York City → is capital of → **USA**

The capital of the USA is Washington, D.C.

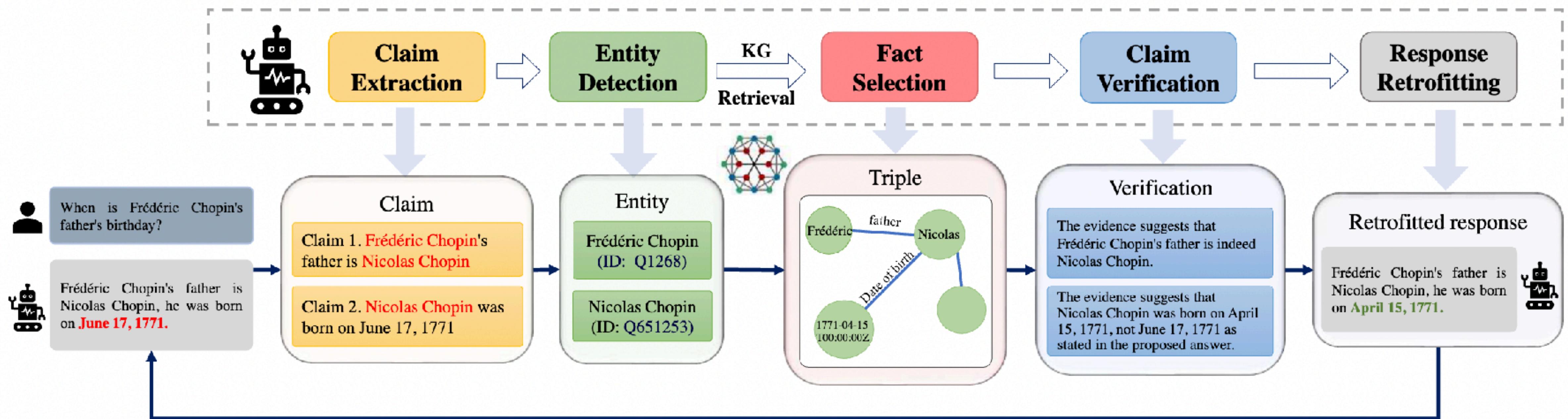


Washington, D.C. → is capital of → **USA**



Retrofitting KGs

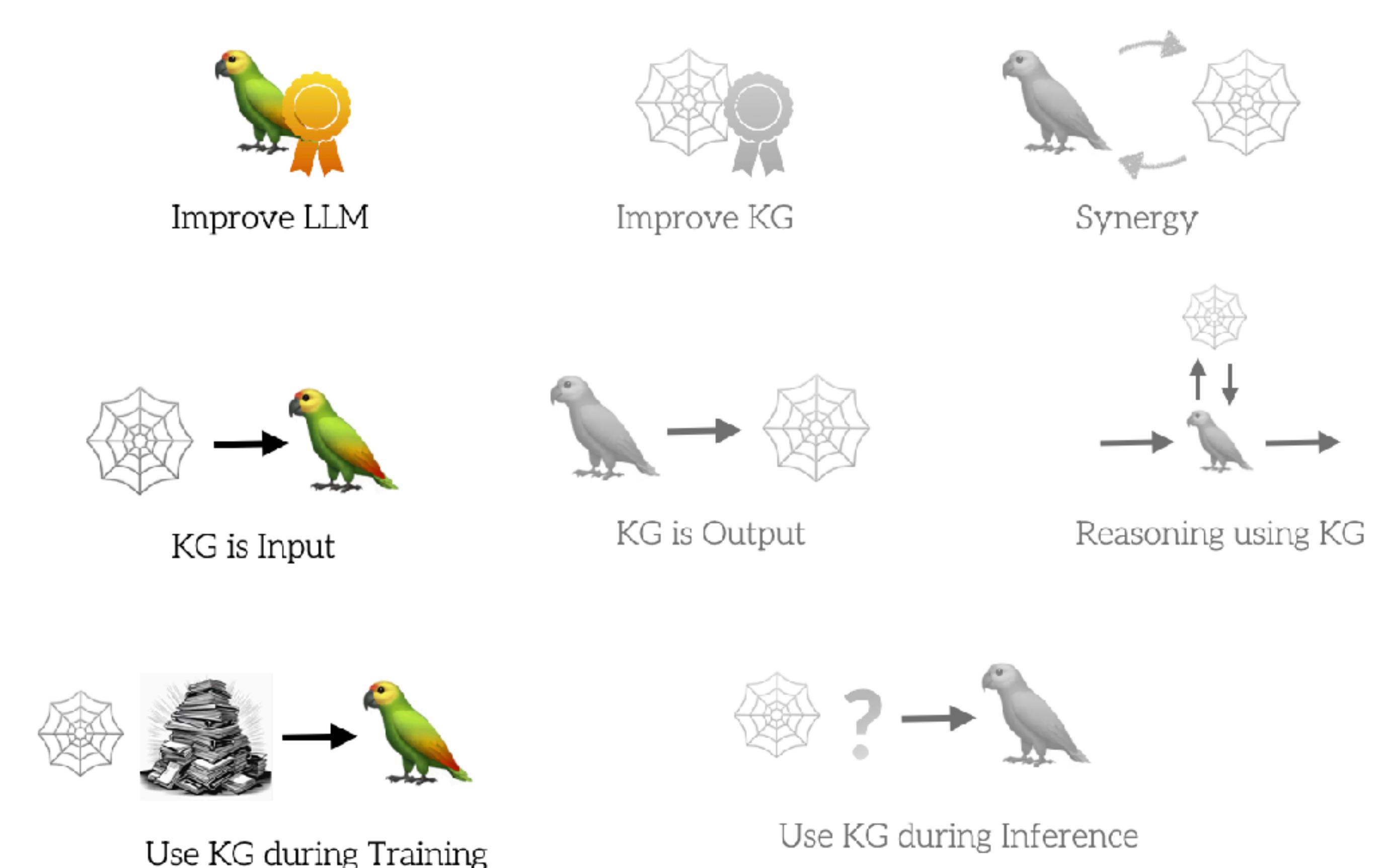
Mitigating Large Language Model Hallucinations via Autonomous Knowledge Graph-based Retrofitting
(Guan et al., 2023)



Training on KGs

SKILL: Structured Knowledge Infusion for Large Language Models (Moiseev, 2022)

- Convert KG into text to use as training data.
- Enhance LLM performance wrt. factual knowledge.



Training on KGs

SKILL: Structured Knowledge Infusion for Large Language Models (Moiseev, 2022)

(Washington, D.C., is capital of, USA)

Option 1: Train directly on triplet

(Washington D.C., [mask], USA)

Option 2: Convert to text

Washington, D.C. is the capital of the USA.

Both methods work well.

KG as Input

- Fine-tuning is often too expensive.



Improve LLM



Improve KG



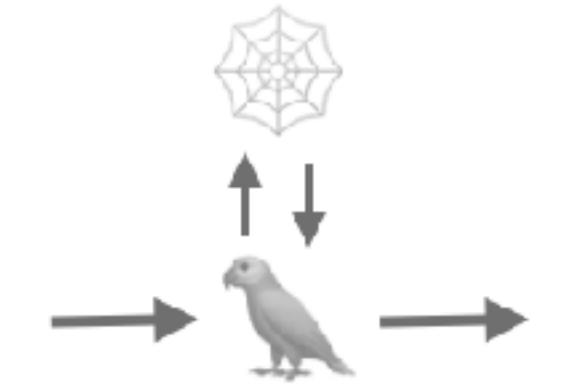
Synergy



KG is Input



KG is Output



Reasoning using KG



Use KG during Training



Use KG during Inference



Part III

Knowledge Graphs for Retrieval Augmented Generation

RAG Recap

*How much money did I spend
on coffee last week?*



Idk.

RAG Recap

How much money did I spend on coffee last week?



Idk.

How much money did I spend on coffee last week?



*Given these financial records:
(3.50 EUR, Cappuccino, 12.08.2024)
(2.80 EUR, Espresso, 15.08.2024)
(2.30 EUR, Cortado, 19.08.2024)*



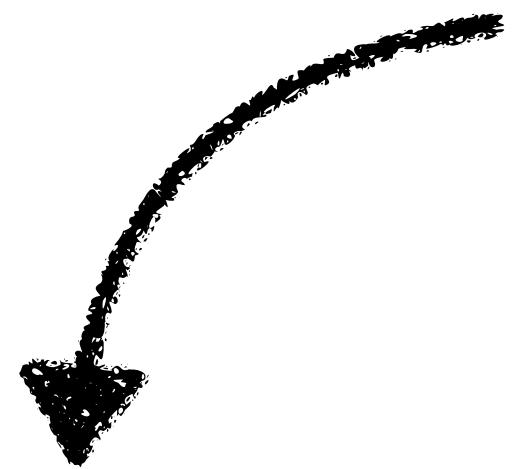
8,60 EUR.

*Answer the question:
How much money did I spend on coffee last week?*

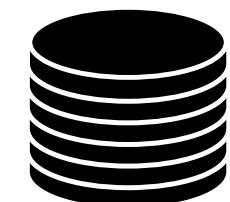
RAG Recap

- We assume there is a corresponding dataset.
 - ▶ private,
 - ▶ up-to-date
 - ▶ specialized
- Challenge: Identify the relevant entries.

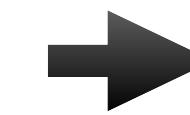
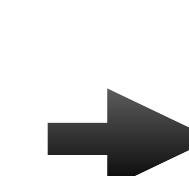
How much money did I spend on coffee last week?



(899.99 EUR	New TV	07.08.2024)
(45.00 EUR	Plants	10.08.2024)
(25.50 EUR	Books	02.08.2024)
(60.00 EUR	Clothing	08.08.2024)
(29.99 EUR	Kitchen Utensils	03.08.2024)
(399.99 EUR	Smartphone	09.08.2024)
(15.00 EUR	Stationery	04.08.2024)
(120.00 EUR	Shoes	06.08.2024)
(300.00 EUR	Furniture	12.08.2024)
(75.00 EUR	Gardening Tools	11.08.2024)
...



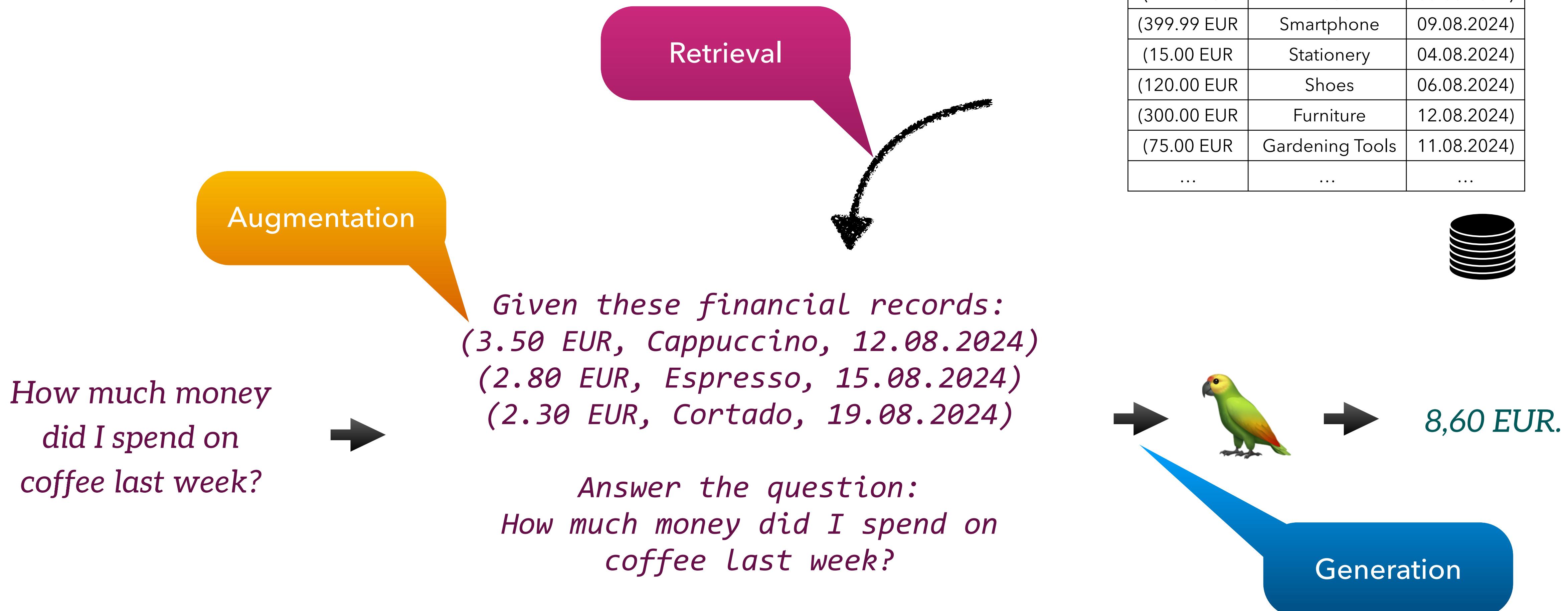
Given these financial records:
(3.50 EUR, Cappuccino, 12.08.2024)
(2.80 EUR, Espresso, 15.08.2024)
(2.30 EUR, Cortado, 19.08.2024)



8,60 EUR.

Answer the question:
How much money did I spend on coffee last week?

RAG Recap



RAG Recap

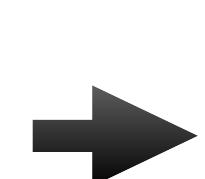
How much money
did I spend on
coffee last week?



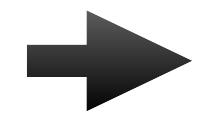
Dataset can also be a KG.



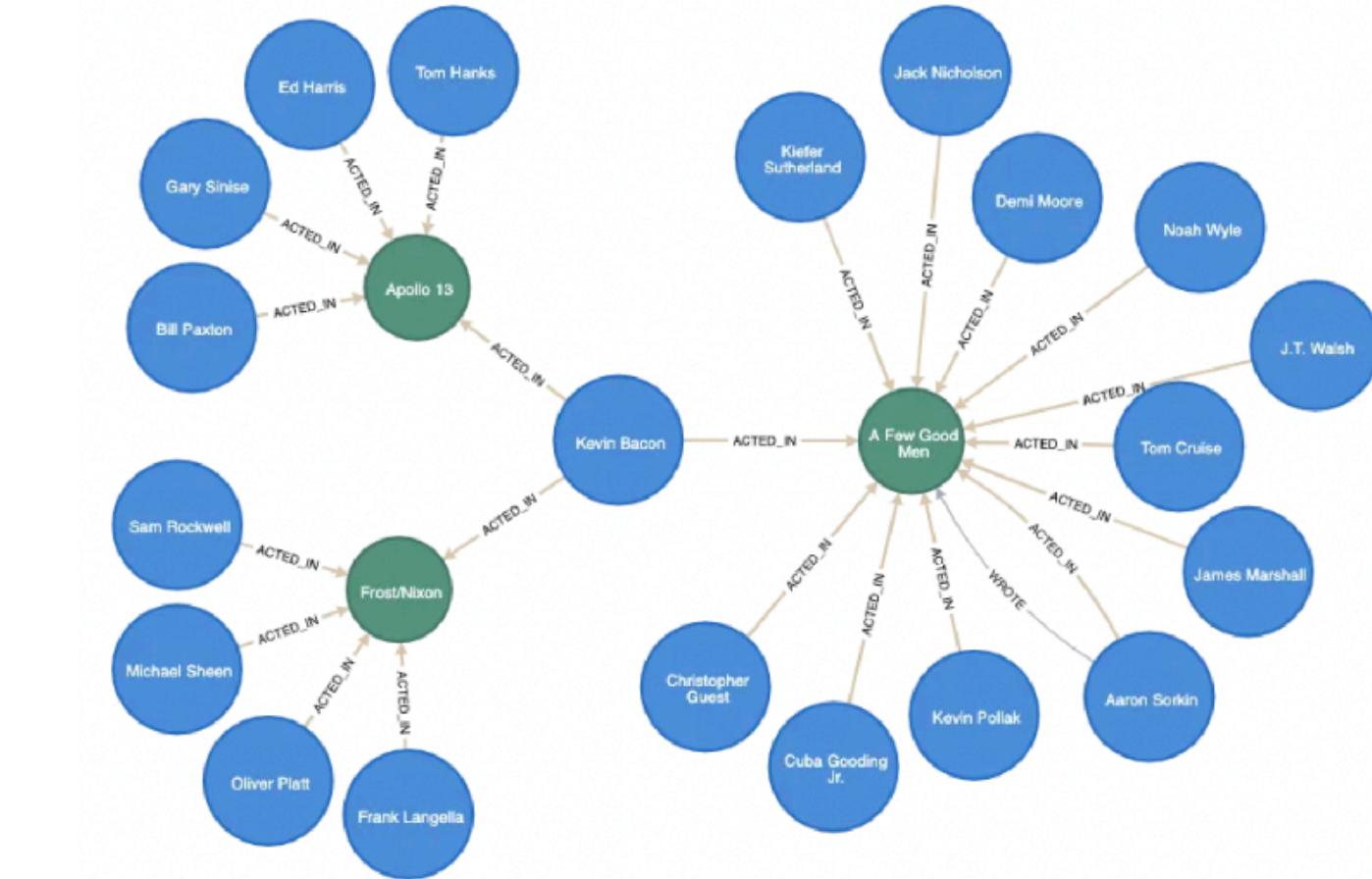
Given these financial records:
(3.50 EUR, Cappuccino, 12.08.2024)
(2.80 EUR, Espresso, 15.08.2024)
(2.30 EUR, Cortado, 19.08.2024)



Answer the question:
*How much money did I spend on
coffee Last week?*



8,60 EUR.



RAG Recap

How much money
did I spend on
coffee last week?



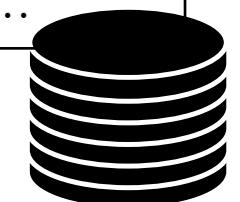
Challenge: Identify the
relevant entries.



Given these financial records:
(3.50 EUR, Cappuccino, 12.08.2024)
(2.80 EUR, Espresso, 15.08.2024)
(2.30 EUR, Cortado, 19.08.2024)

Answer the question:
*How much money did I spend on
coffee last week?*

(899.99 EUR	New TV	07.08.2024)
(45.00 EUR	Plants	10.08.2024)
(25.50 EUR	Books	02.08.2024)
(60.00 EUR	Clothing	08.08.2024)
(29.99 EUR	Kitchen Utensils	03.08.2024)
(399.99 EUR	Smartphone	09.08.2024)
(15.00 EUR	Stationery	04.08.2024)
(120.00 EUR	Shoes	06.08.2024)
(300.00 EUR	Furniture	12.08.2024)
(75.00 EUR	Gardening Tools	11.08.2024)
...



8,60 EUR.

RAG Recap

1) Embedding

acting as

Vector Index



How much money
did I spend on
coffee last week?



Given these financial records:
(3.50 EUR, Cappuccino, 12.08.2024)
(2.80 EUR, Espresso, 15.08.2024)
(2.30 EUR, Cortado, 19.08.2024)



8,60 EUR.

*Answer the question:
How much money did I spend on
coffee last week?*

RAG Recap

2) Identify entries

with **similar**

embeddings in **vector**

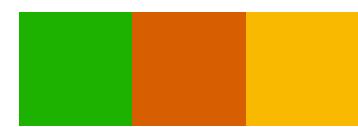
database.



Gardening Tools



Espresso



Shoes



Cappuccino



New TV



Plants



Cortado

1) Embedding

acting as

Vector Index



How much money
did I spend on
coffee last week?



Given these financial records:
(3.50 EUR, Cappuccino, 12.08.2024)
(2.80 EUR, Espresso, 15.08.2024)
(2.30 EUR, Cortado, 19.08.2024)

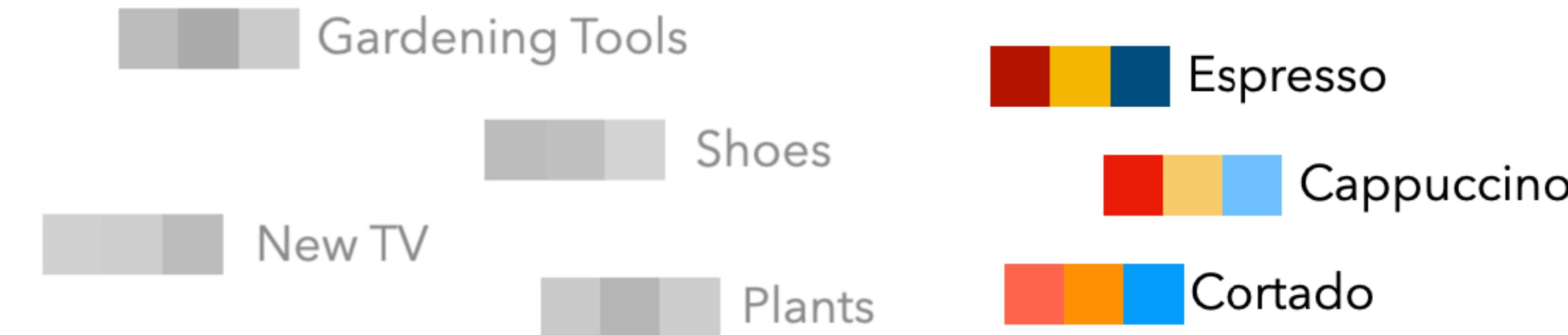


8,60 EUR.

*Answer the question:
How much money did I spend on
coffee last week?*

RAG Recap

2) Identify entries
with **similar**
embeddings in **vector**
database.



1) Embedding
acting as
Vector Index



How much money
did I spend on
coffee last week?



Given these financial records:
(3.50 EUR, Cappuccino, 12.08.2024)
(2.80 EUR, Espresso, 15.08.2024)
(2.30 EUR, Cortado, 19.08.2024)

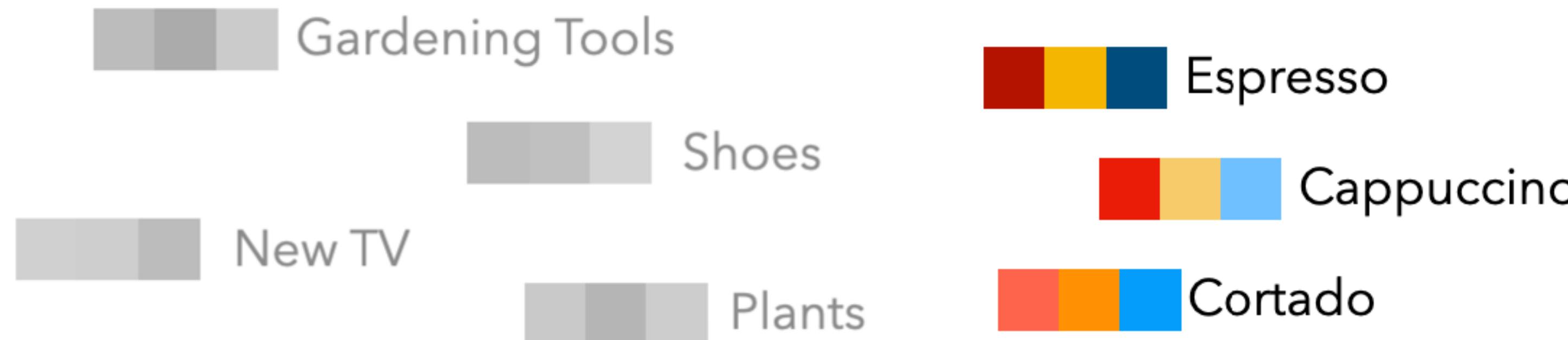


8,60 EUR.

Answer the question:
*How much money did I spend on
coffee last week?*

RAG Recap

2) Identify entries
with **similar**
embeddings in **vector**
database.



1) **Embedding**
acting as
Vector Index



How much money
did I spend on
coffee last week?



Given these financial records:
(3.50 EUR, Cappuccino, 12.08.2024)
(2.80 EUR, Espresso, 15.08.2024)
(2.30 EUR, Cortado, 19.08.2024)



3) **Add** these to
the prompts

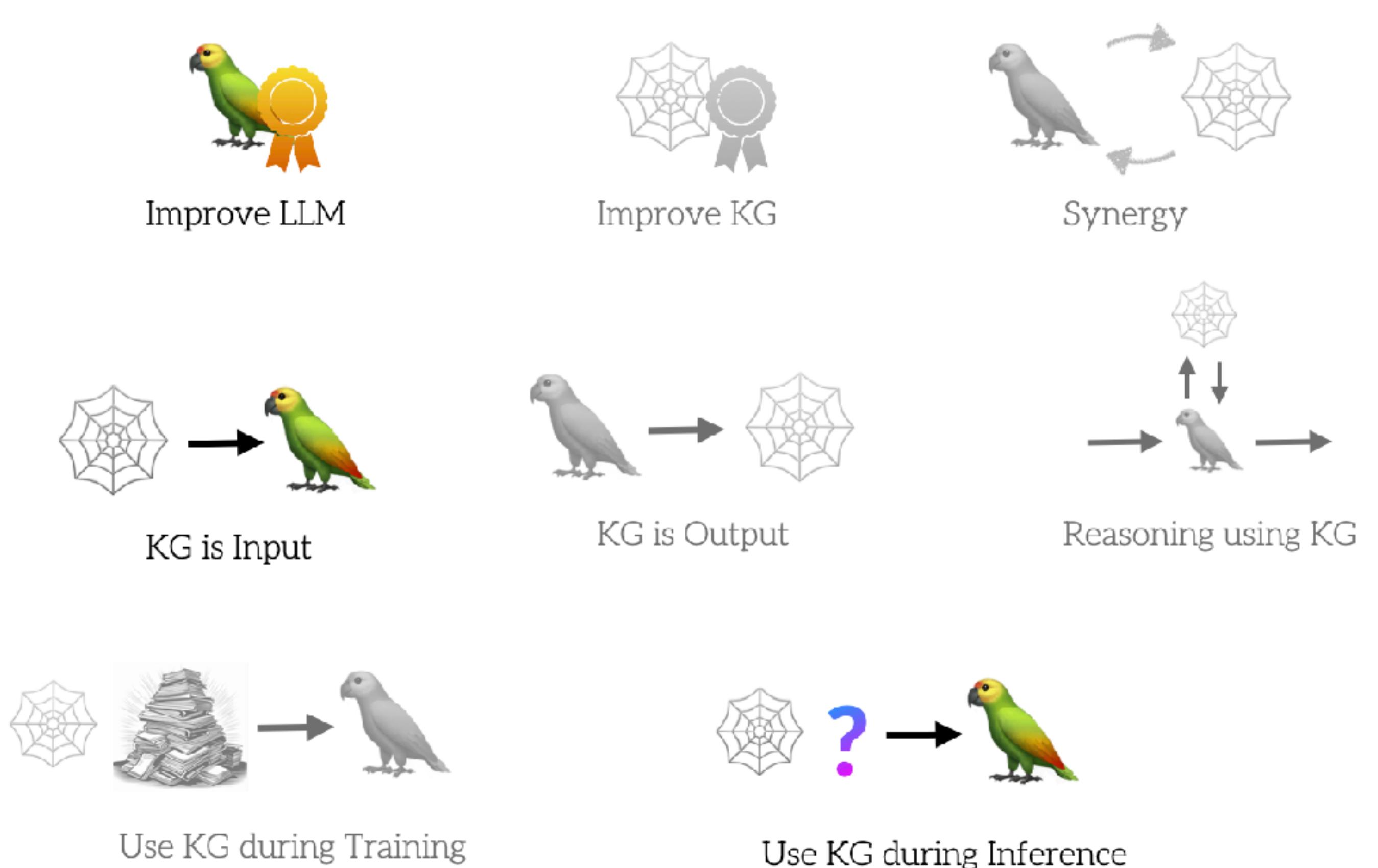
8,60 EUR.

Answer the question:
How much money did I spend on
coffee last week?

KG-Based RAG

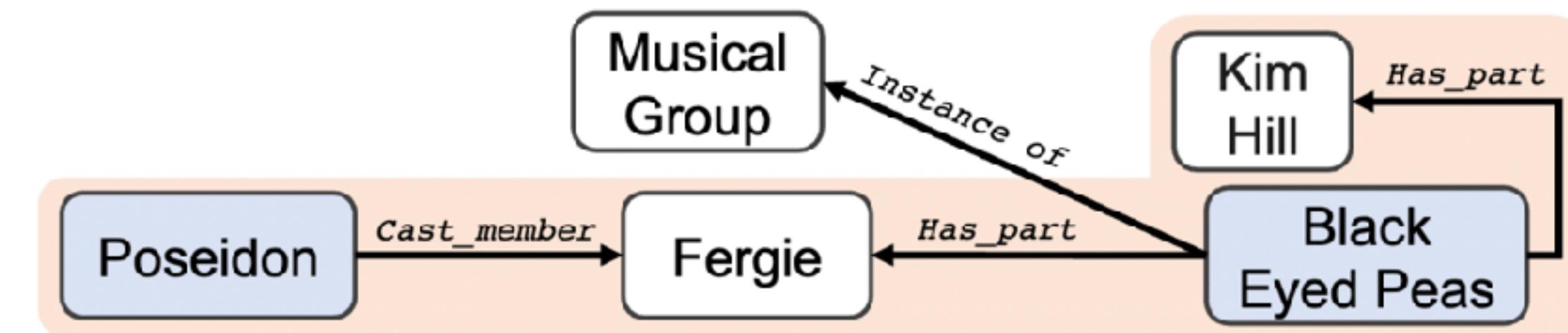
Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering
(Baek et al., 2023)

- Augment the input of LLMs with relevant facts retrieved from a knowledge graph.



KG-Based RAG

Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering
(Baek et al., 2023)

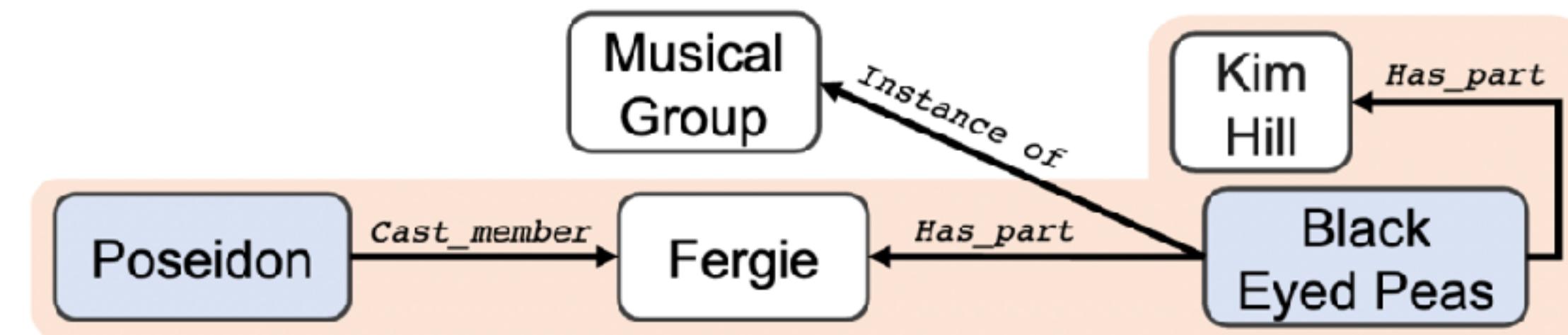


*Which member of Black
Eyed Peas appeared in
Poseidon?*



KG-Based RAG

Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering
(Baek et al., 2023)



Below are the facts that might be relevant to answer the question:

*(Black Eyed Peas, has part, Fergie),
(Black Eyed Peas, has part, Kim Hill),
(Poseidon, cast member, Fergie)*

*Question:
Which member of Black Eyed Peas appeared in Poseidon? Answer:*

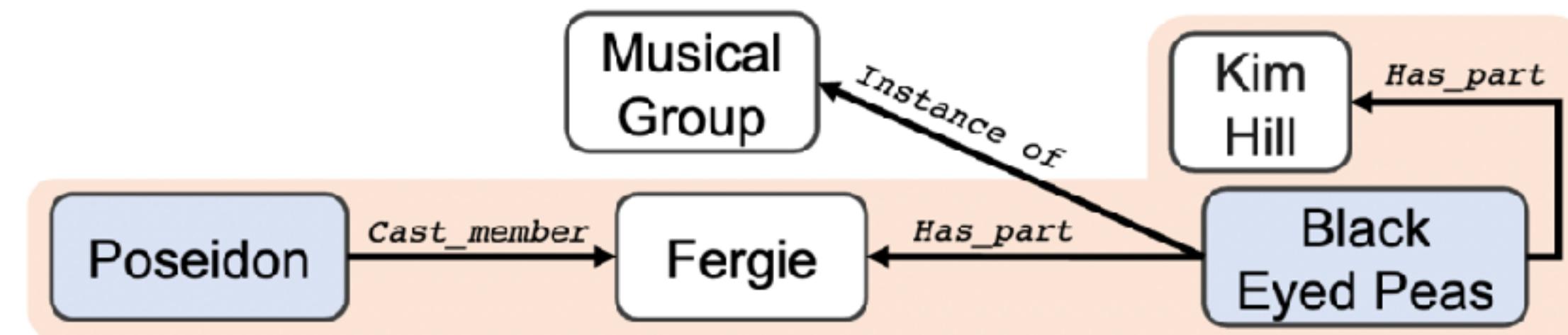
Which member of Black Eyed Peas appeared in Poseidon?



Fergie

KG-Based RAG

Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering
(Baek et al., 2023)



Entity extraction/matching
+ embedding similarity

Which member of Black Eyed Peas appeared in Poseidon? →

Below are the facts that might be relevant to answer the question:

*(Black Eyed Peas, has part, Fergie),
(Black Eyed Peas, has part, Kim Hill),
(Poseidon, cast member, Fergie)*

*Question:
Which member of Black Eyed Peas appeared in Poseidon? Answer:*

→ *Fergie*

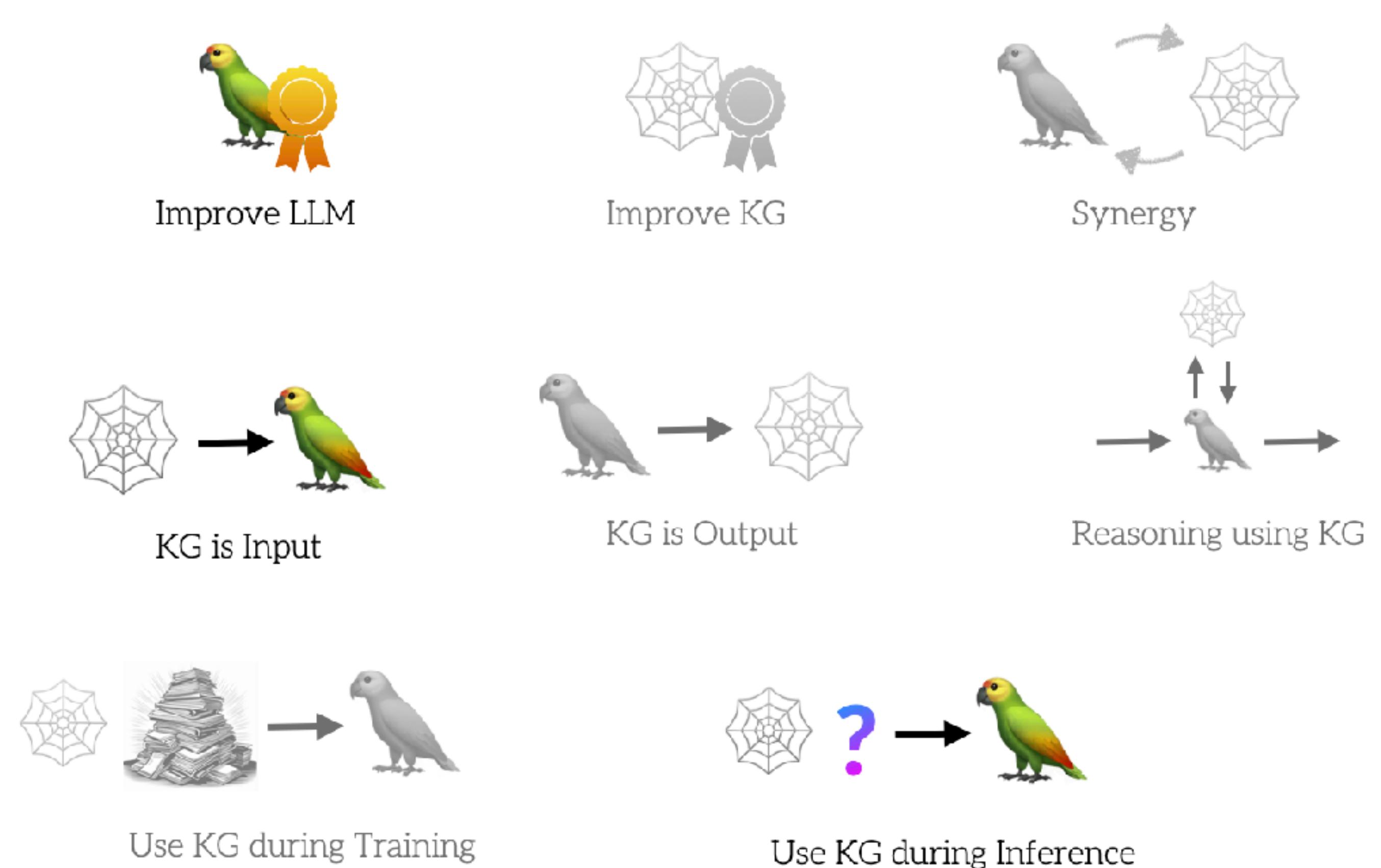
KG-Based RAG

From Local to Global: A Graph RAG Approach to Query-Focused Summarization (Edge et al., 2024)

- **(Vanilla) RAG struggles with global questions:**

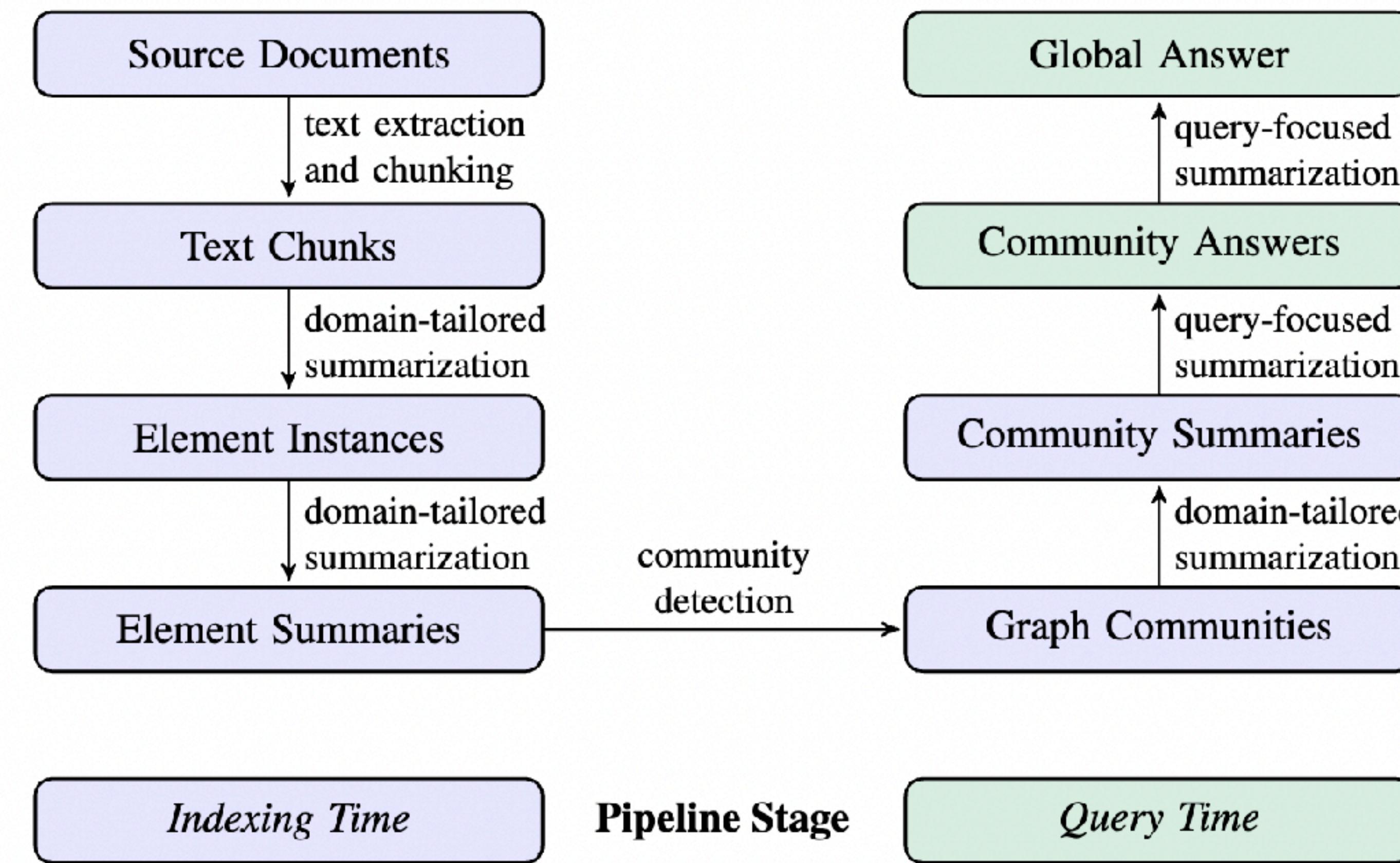
„What are the main themes in the dataset?“

- **Convert dataset to KG + exploit graph structure for retrieval.**



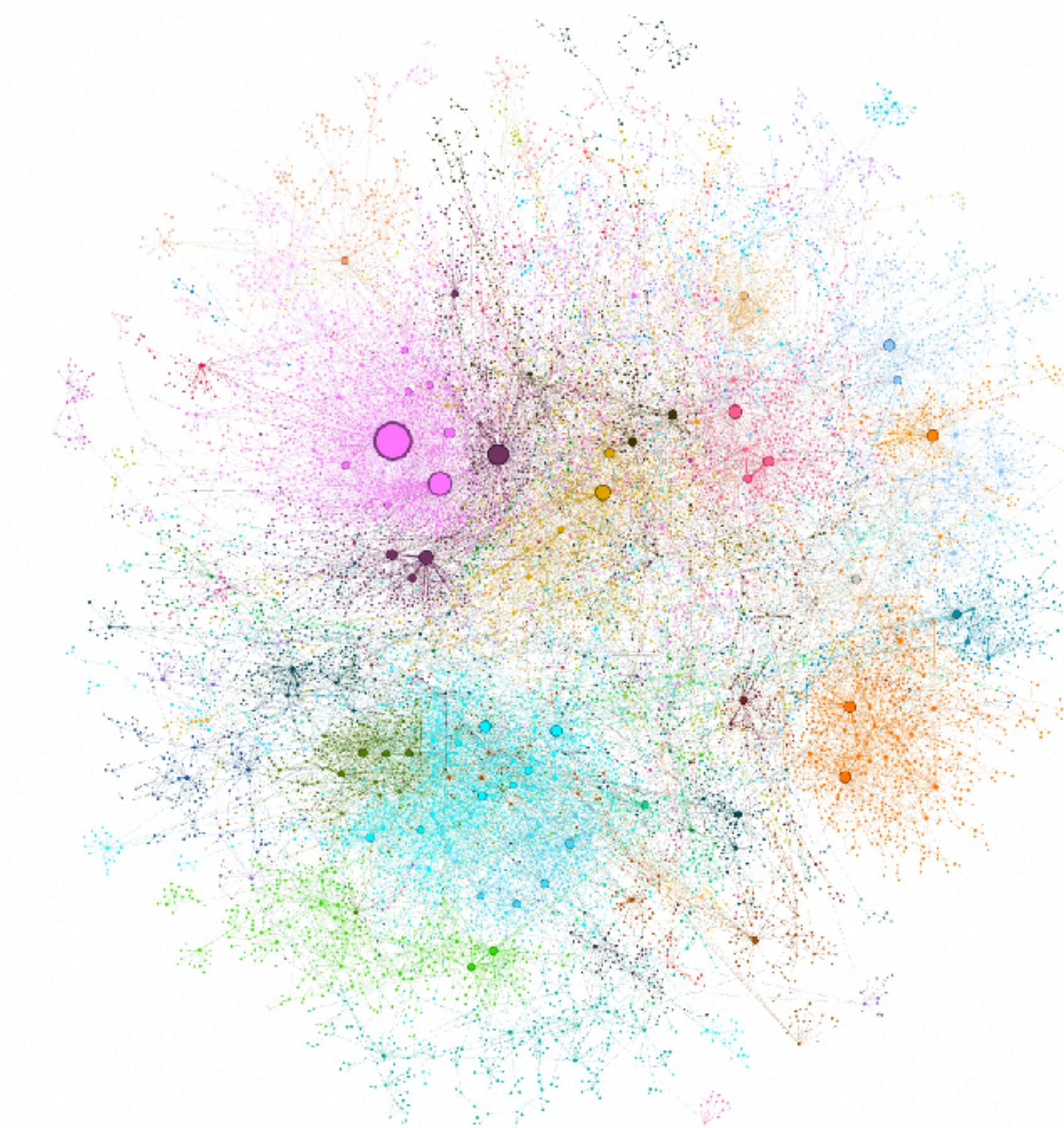
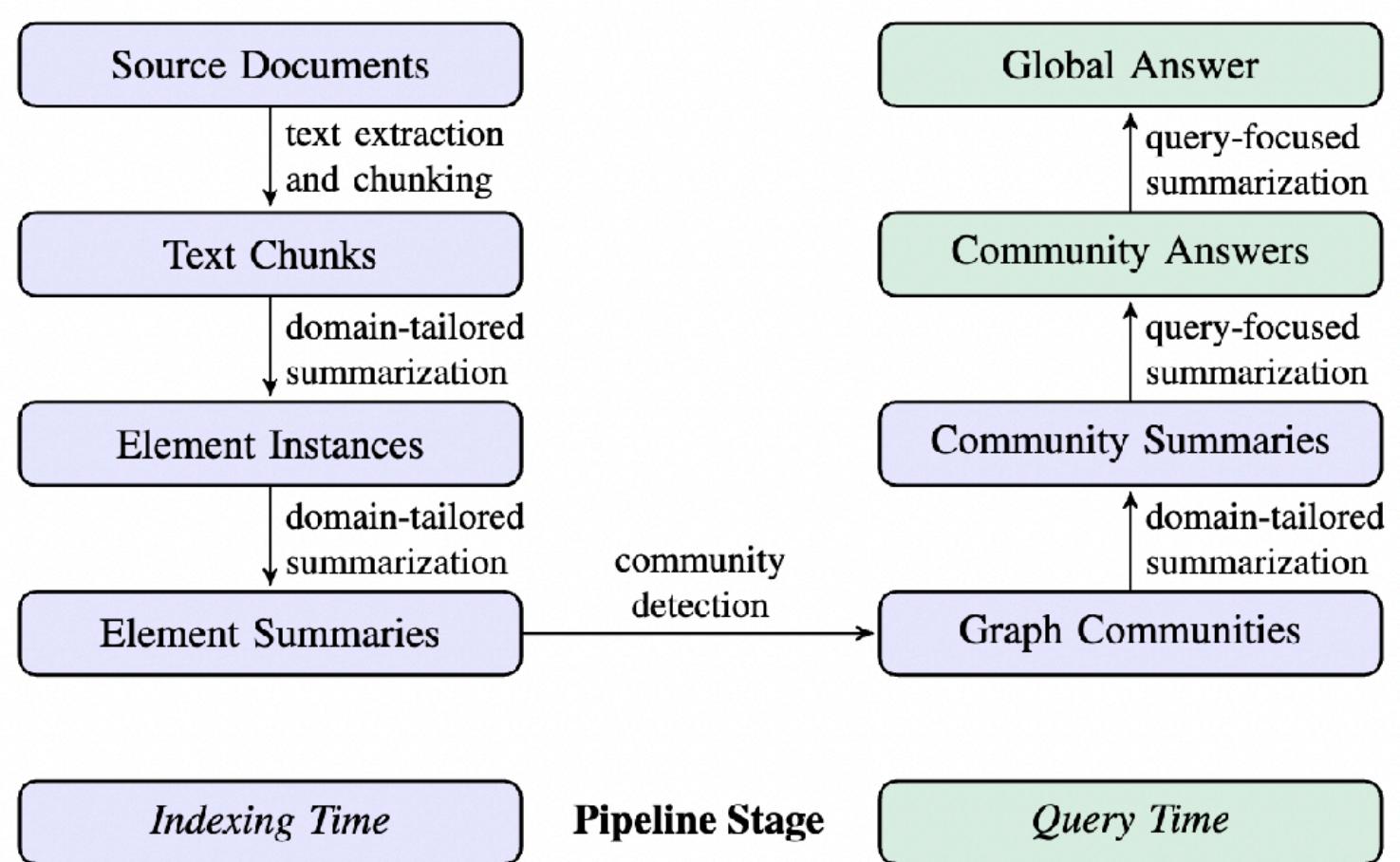
KG-Based RAG

From Local to Global: A Graph RAG Approach to Query-Focused Summarization (Edge et al., 2024)

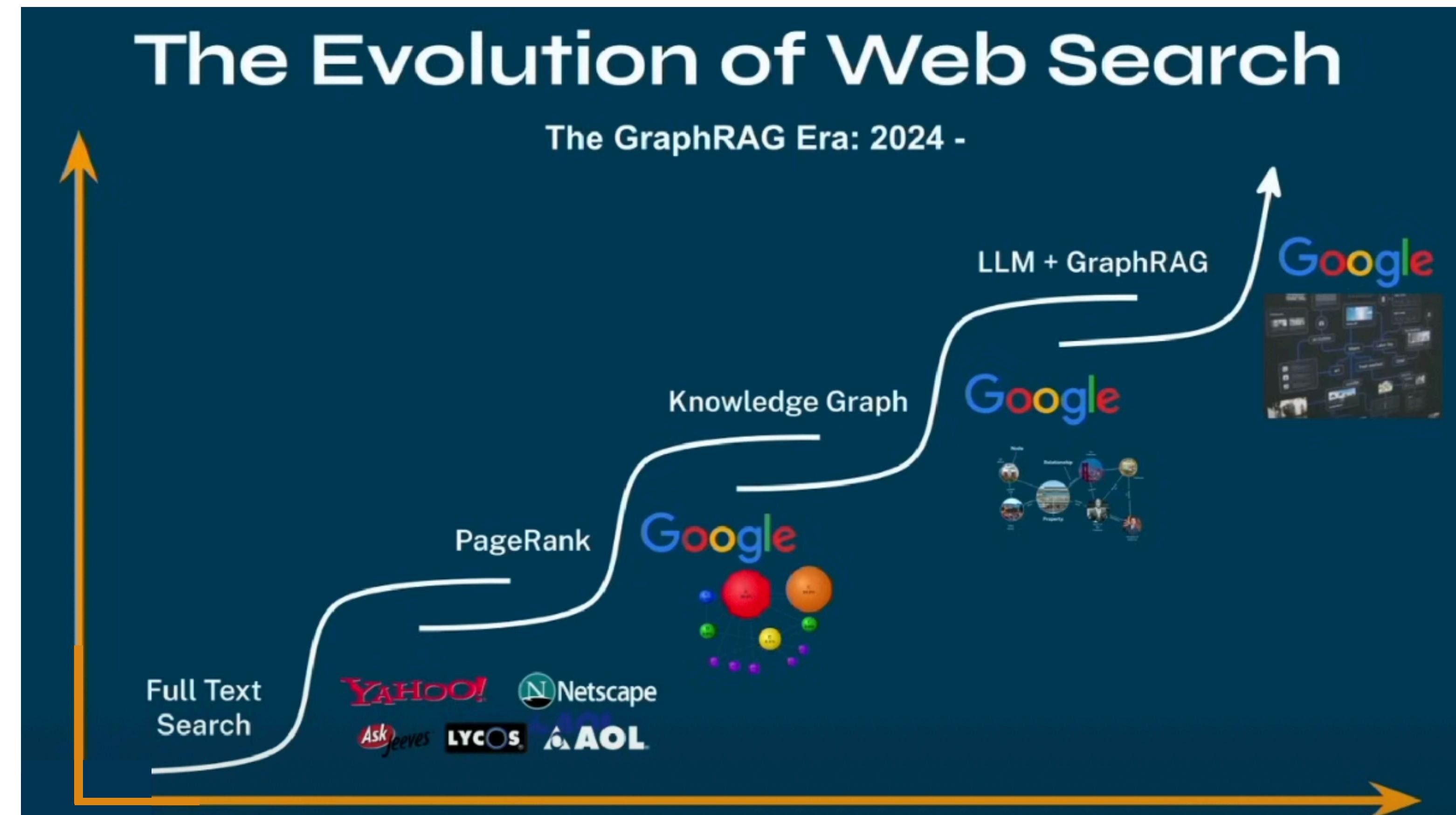


KG-Based RAG

From Local to Global: A Graph RAG Approach to Query-Focused Summarization (Edge et al., 2024)



Outlook



GraphRAG: The Marriage of Knowledge Graphs and RAG: Emil Eifrem

(<https://youtu.be/knDDGYHnnSI?si=apd0K6ZZ7jYtuq84>)

Build Your Own GraphRAG



Ollama



LitGPT



OpenAI



Amazon



HuggingFace



Neo4j



NebulaGraph



LangChain



Llamaindex

Main LLM

Embedding LLM

Graph database

Toolchain

Build Your Own GraphRAG



Ollama



LitGPT



OpenAI



Amazon



HuggingFace



Neo4j



NebulaGraph



LangChain



LlamalIndex

Main LLM

Embedding LLM

Graph database

Toolchain

Generate KG:

- Create your own KG manually
- Neo4J: LLM graph builder
- Microsoft GraphRAG
- LlamalIndex: KnowledgeGraphIndex

Summary



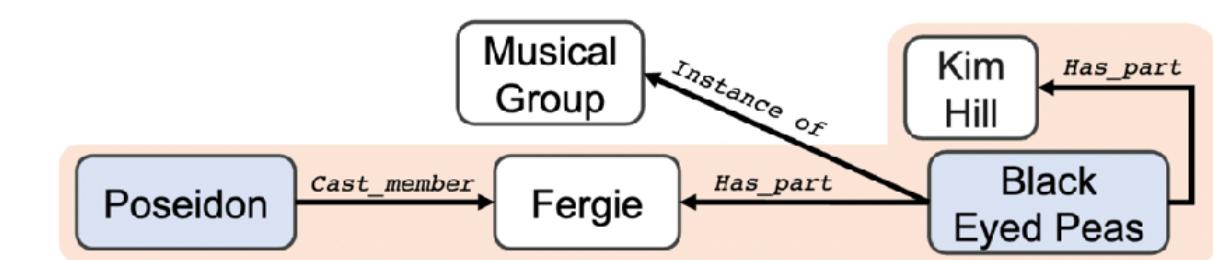
Everything is a graph.



KGs can improve LLMs.
LLMs can improve KGs.



RAG equips LLMs with
real-time and private data.



Graph RAG =
Higher accuracy + global answers

We are hiring: datasciapps.de/jobs
Anonymous feedback: tellonym.me/LLMsummerschool
All references are available on: github.com/gerritgr/LLMSummerSchool



References

Surveys:

- Unifying Large Language Models and Knowledge Graphs: A Roadmap (Pan et al., 2023)
- Graph Machine Learning in the Era of Large Language Models (Fan et al., 2024)

Tutorials:

- ArangoDB - LLMS With Knowledge Graph (youtube.com/watch?v=DkbX8O9zd_8)
- AI Engineer World's Fair 2024 (talk by Emil Eifrem highly recommended)
- Neo4j Graphacademy
- ACL 2023 Tutorial: Complex Reasoning in Natural Language (wenting-zhao.github.io/complex-reasoning-tutorial/)

We are hiring: datasciapps.de/jobs

Anonymous feedback: tellonym.me/LLMsummerschool

All references are available on: github.com/gerritgr/LLMSummerSchool

