

Hallucinations

Reducing Hallucinations in ChatGPT like systems

Abhimanyu Aryan

FullStack Software Engineer,
CEGID*

Voxel51

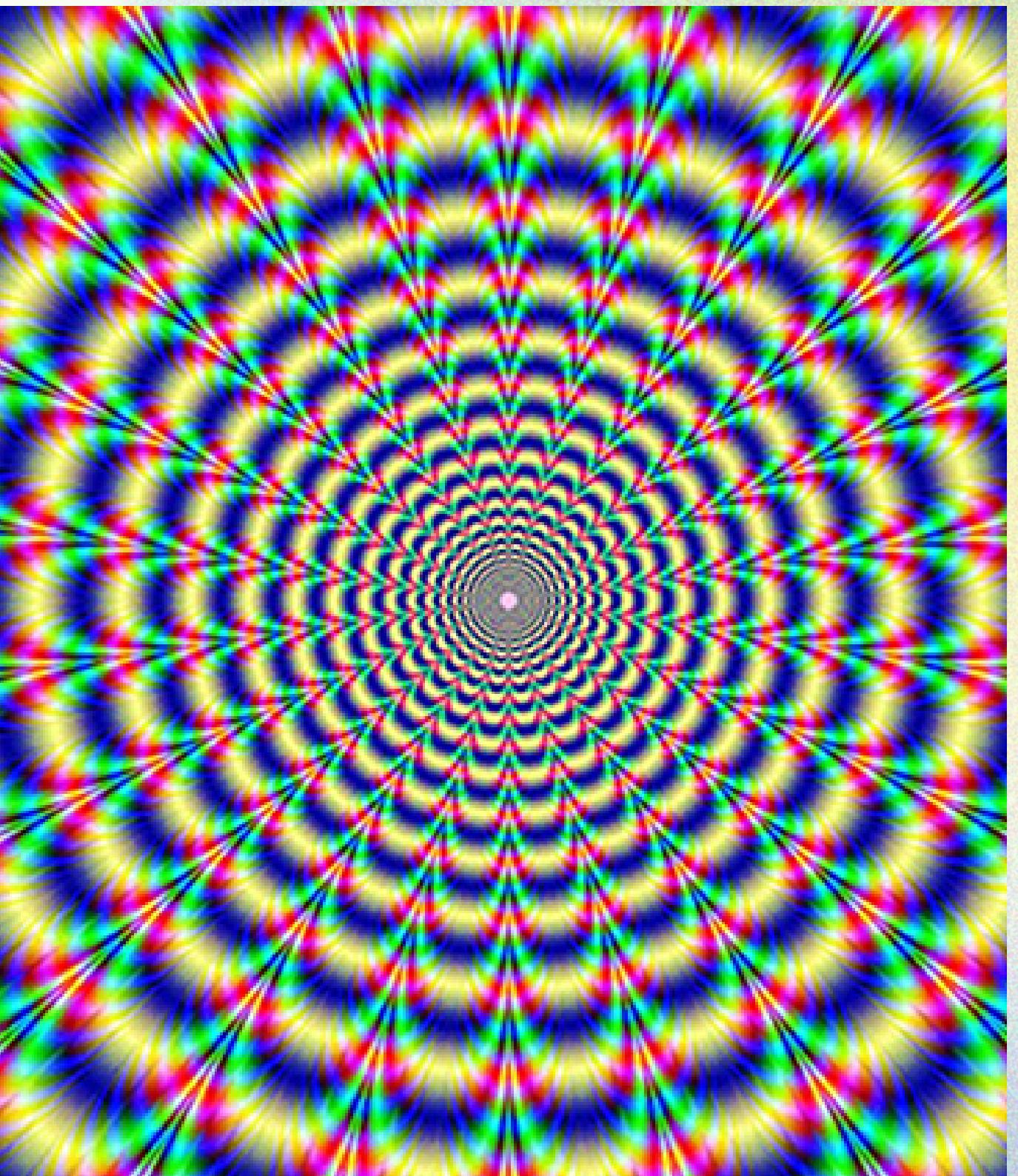
Agenda

- What is Hallucination
- Why LLMs Hallucinate
- Methodology
- Ways KG can Reduce Hallucination
- Why KG?
- Cypher Generation
- Accurate Cypher Generation
- Unstructured Data to KG
- GraphRAG
- Future
- Questions

Hallucination

In context of LLMs, hallucination refers to the generation of content that is incorrect, nonsensical, or not real

- Factual errors
- Inconsistent responses



Why

Temperature

LLMs employ a temperature setting that controls the degree of randomness in their text generation

Missing Information

LLMs are trained on massive datasets that can take months to compile, making it challenging to keep them up-to-date.

Model Training and Complexity

The complexity of Large Language Models, along with the possibility of training on inaccurate or misleading data, can lead to unpredictable or incorrect outputs.

Methodology

Improving LLM accuracy

Prompt Engineering	In-Context Learning	Fine-Tuning	Grounding
Art of crafting specific instructions to guide LLMs towards desired outcomes.	In-context learning helps the model better understand a task by supplying examples that guide its responses.	Fine-tuning refers to further training a language model on a smaller, task-specific dataset after its initial training.	Grounding enables a language model to access external, up-to-date sources or databases to improve the accuracy of its responses.

Knowledge Graph

Nodes

Objects

Edges

Events

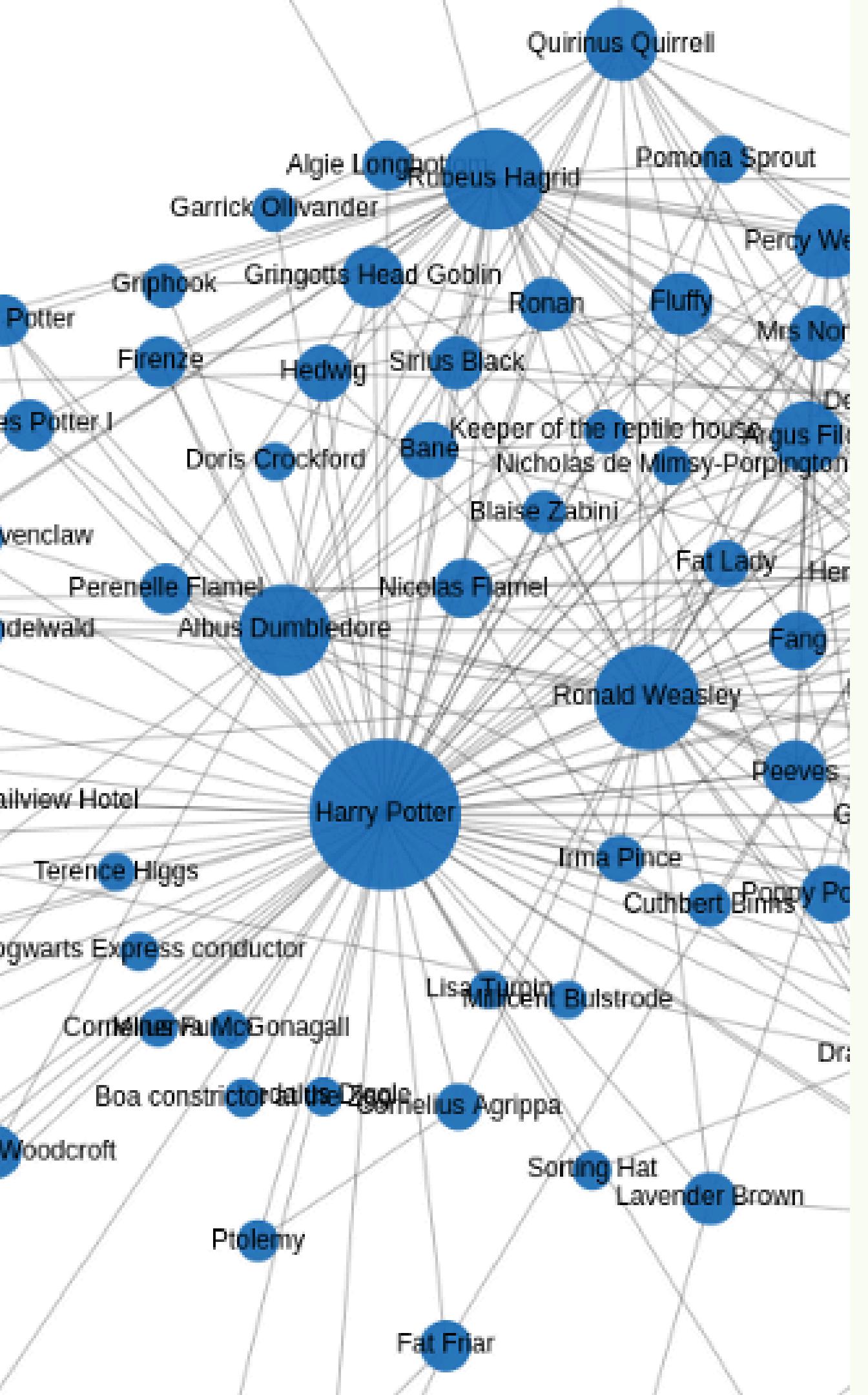
Labels

Situations

Ontologies

Concepts

What is a knowledge Graph?



How Knowledge Graphs Reduce Hallucinations

Providing Factual Grounding

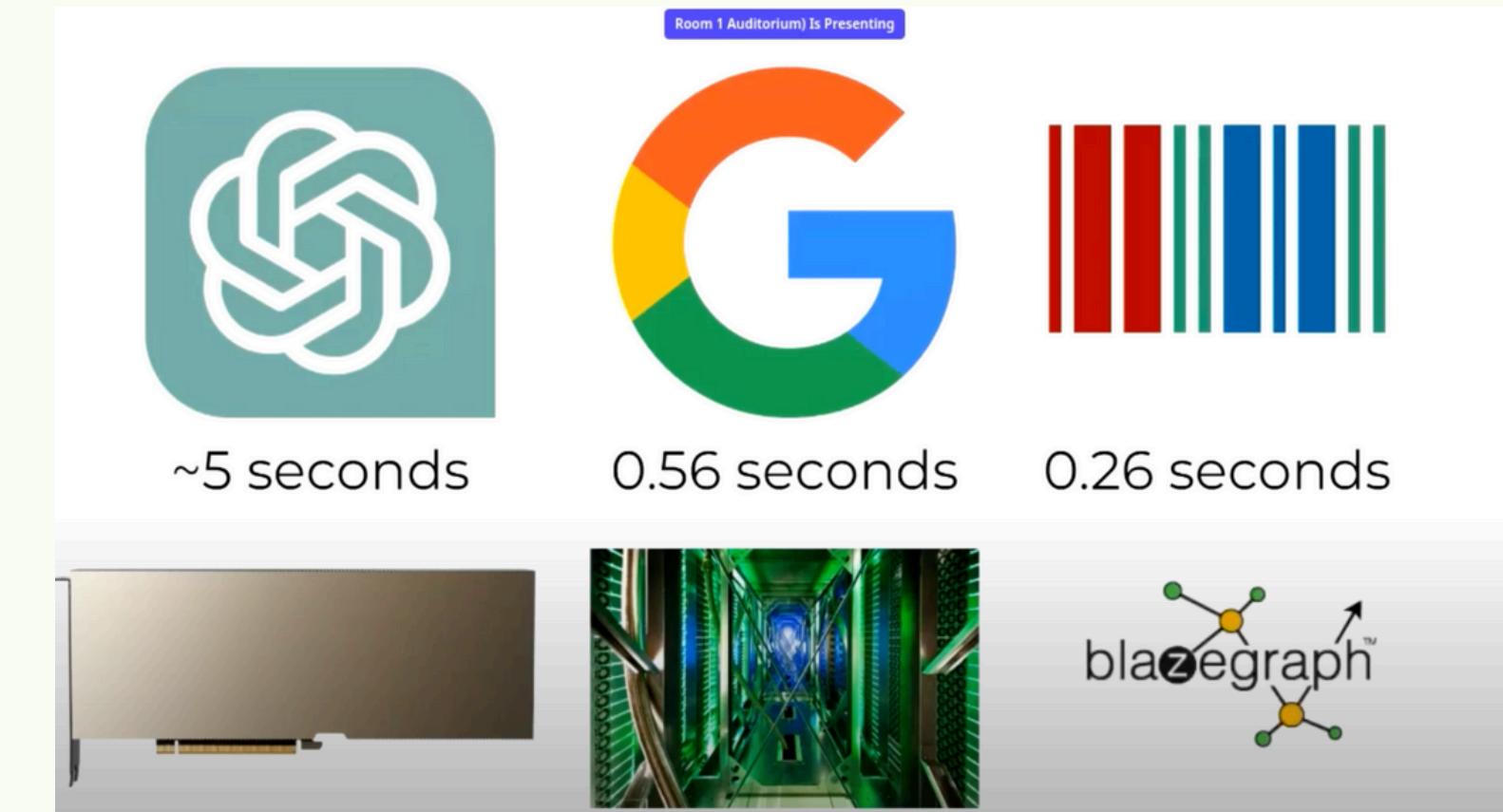
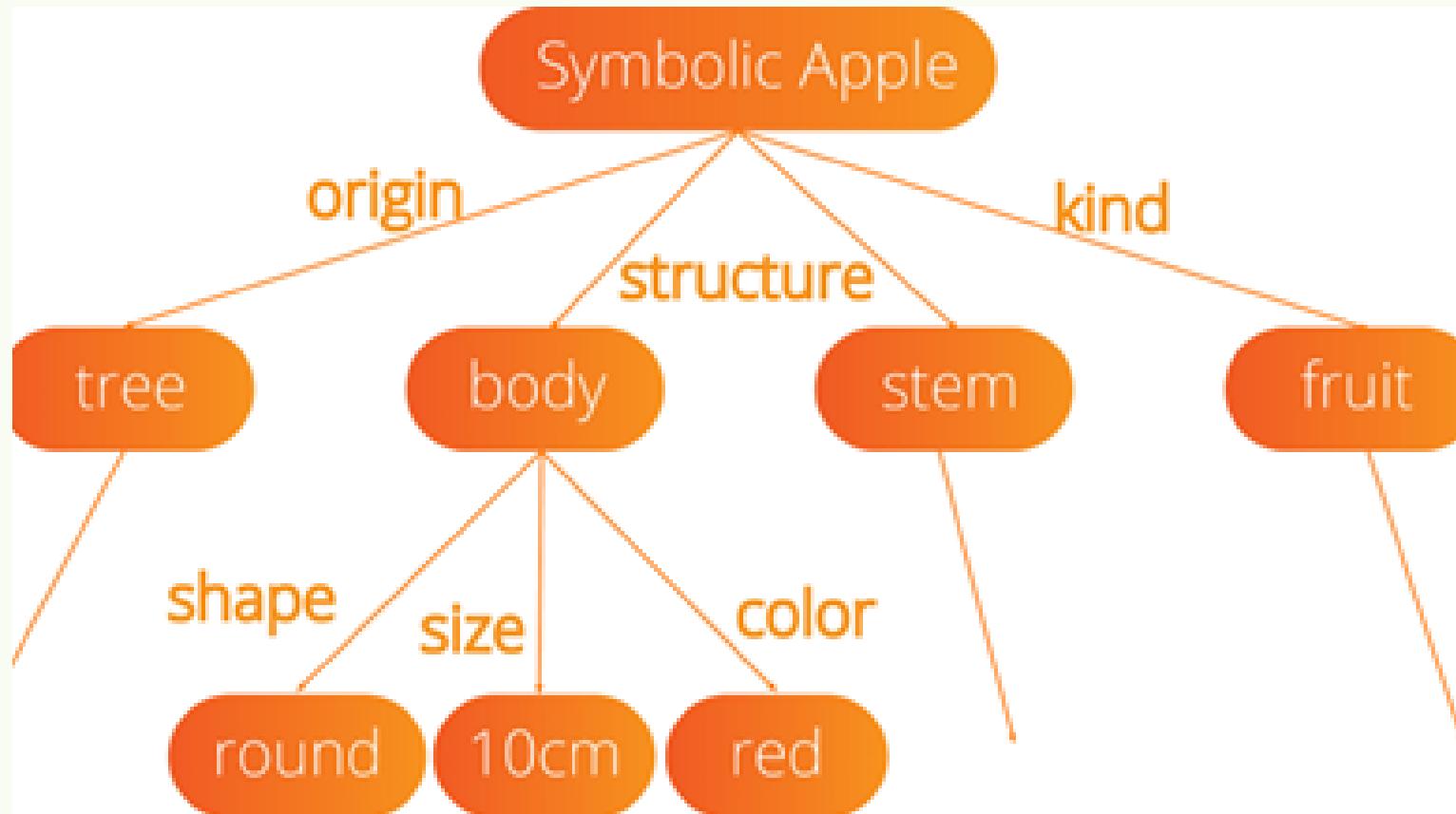
Improving Reasoning
Capabilities

Enhancing Retrieval

Reducing Over-reliance on
Statistical Patterns

Why KG?

source: [KGC23 Keynote](#)



LLMs Hallucinate

A key to cracking the hallucinations problem is adding knowledge graphs to vector-based retrieval augmented generation (RAG)

LLMs are slow and expensive

Wikimedia query retrieval on blazegraph is magnitudes faster than competition running on legacy hardware

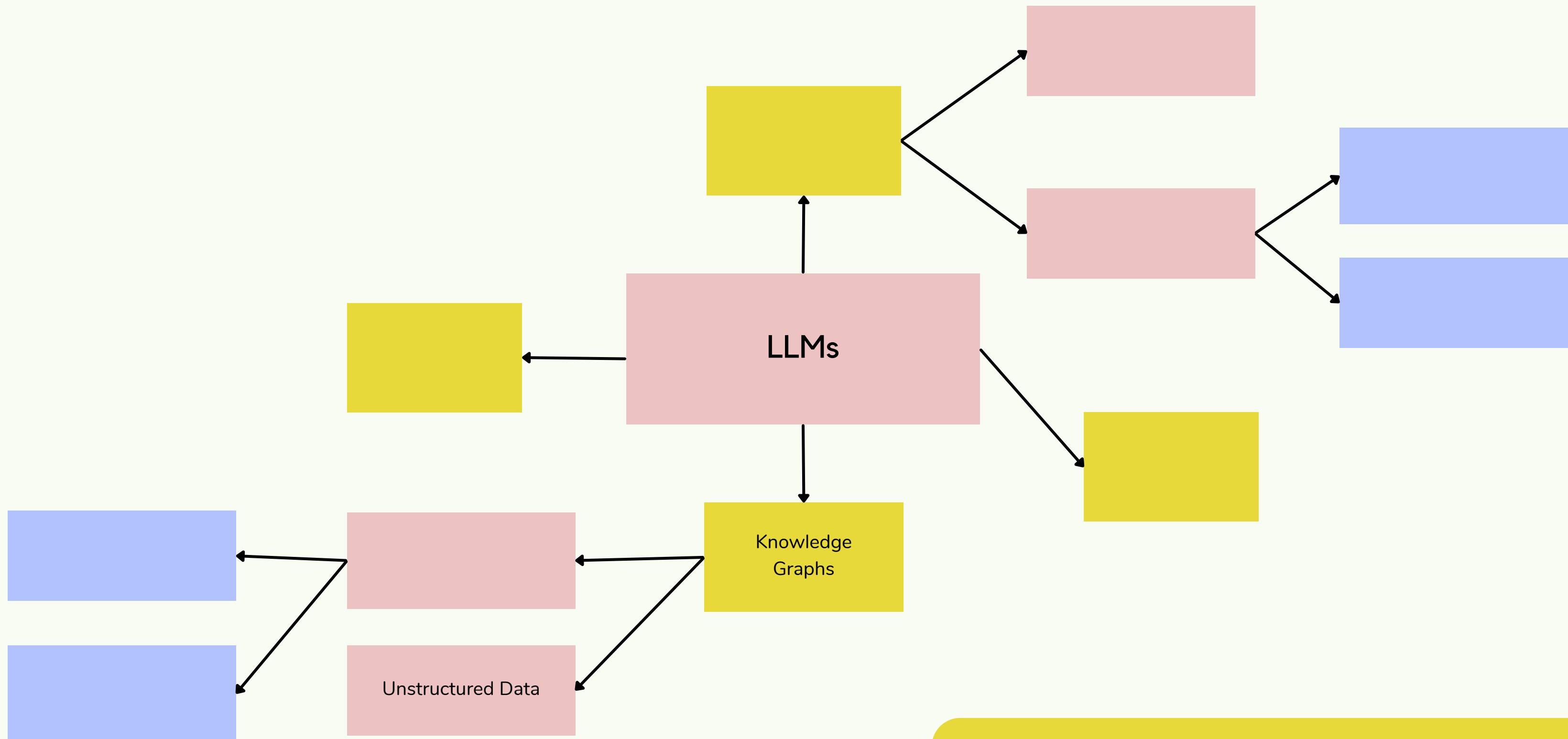
Generation Cypher with LLMs

The LLM doesn't return consistent results - its objective is to produce an answer, not the same response. The response may not be correct or even generate an error due to invalid Cypher.

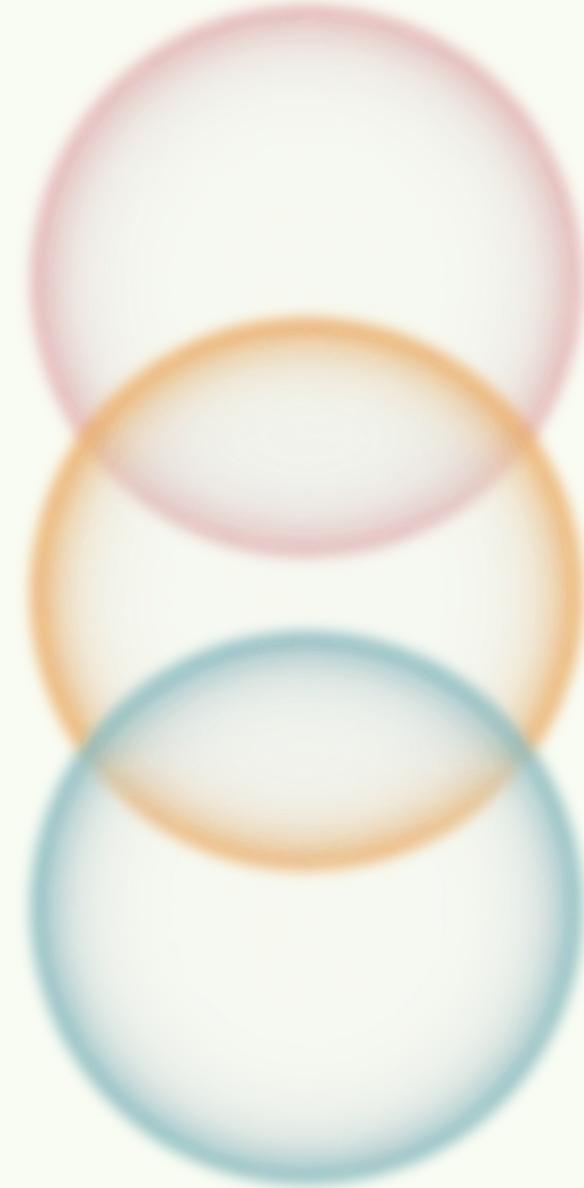
Methodology to fix cypher

Template Instructions	Few Shot Prompting
<ul style="list-style-type: none">• Provide specific instructions to the LLM to state that the generated Cypher statements should follow the schema.• Provide relationship types and properties in the schema.	<ul style="list-style-type: none">• A technique where you provide the LLM with an example of how to respond or generate a response to a specific scenario.

Generating KG from Unstructured Data



Project NaLLM: Exploring and demonstrating the synergies
between Neo4j and Large Language Models (LLMs)

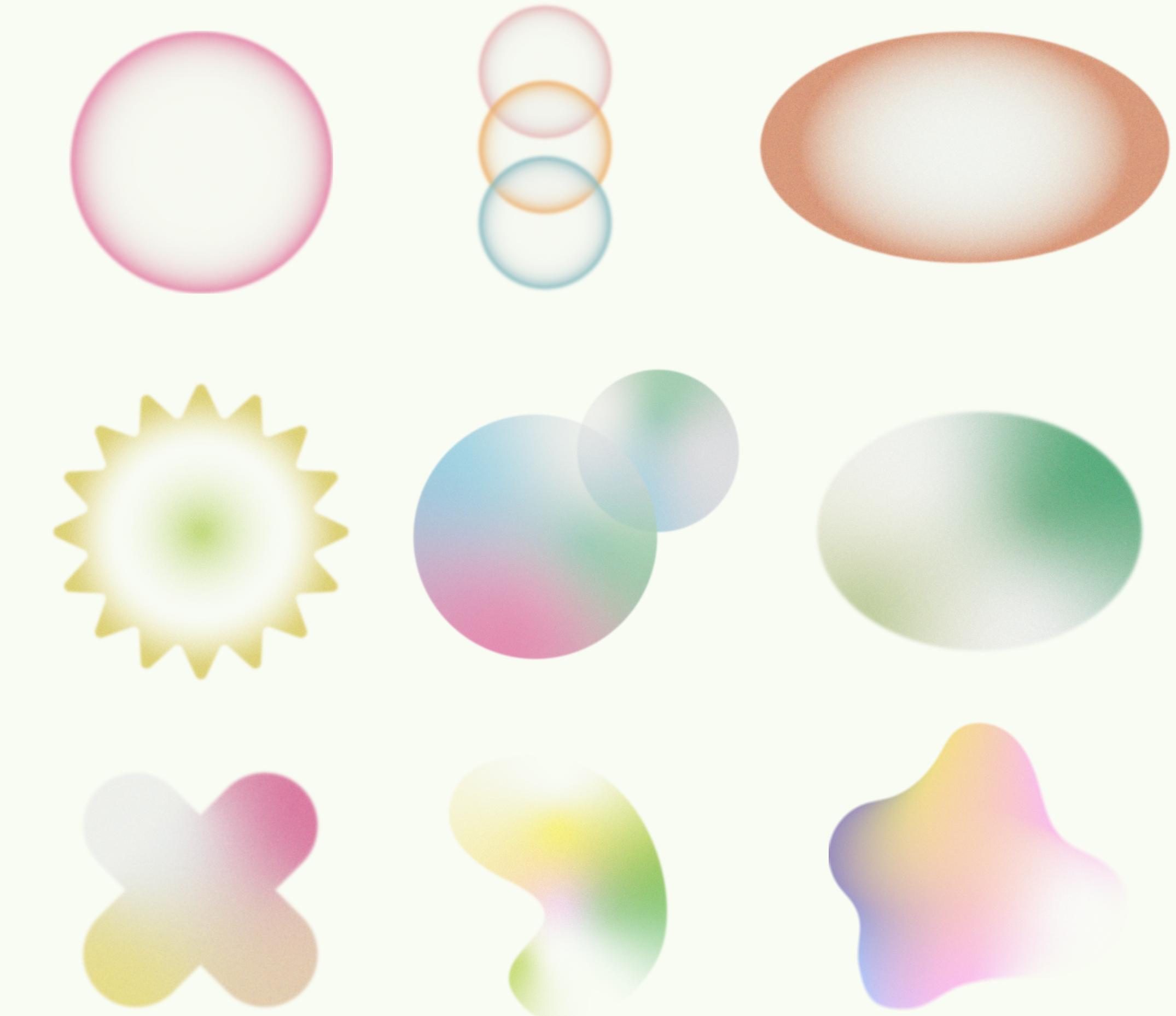


GraphRAG

Unlocking LLM discovery on narrative private data

Future

Knowledge graph, bright and clear,
Connecting nodes, far and near. Data
linked, a cosmic view, Learning new
things, me and you





Thank you!

Do you have any questions?