

# Knowledge Extraction from Databases

**João Caldeira**

Assistant Professor

Email. [joao.caldeira@ulusofona.pt](mailto:joao.caldeira@ulusofona.pt)

March 19, 2025



# Knowledge Extraction from Databases

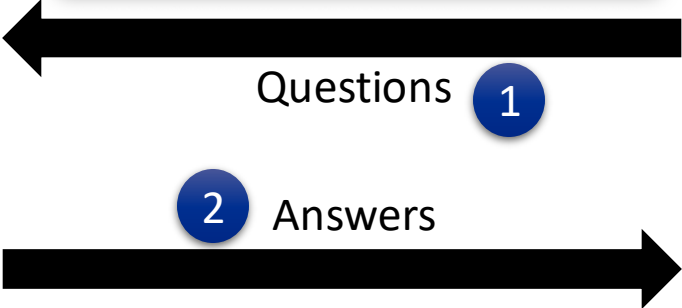
## Architecture

Models (gpt4, etc...)



Models (llama3.2, etc...)

```
Last login: Fri Mar 7 15:43:32 on tty010
josealberto
complete:13: command not found: compdef
> ollama list
name          ID          SIZE  MODIFIED
all-minima:latest 1b226e2802db 45 MB  2 months ago
nomic-embed-text:latest 6a189f422b67 274 MB 2 months ago
mxbai-embed-large:latest 4d88395d2d67 669 MB 2 months ago
mistral:latest    f974a74358d6 4.3 GB 4 months ago
phi3.5:latest     61819f5378a3 2.2 GB 4 months ago
qwen2.5:latest    8450da08e4d8 6.7 GB 4 months ago
gemma2:latest     ff02c3702f32 5.4 GB 4 months ago
llava:latest      86d30f6b0cb1 4.7 GB 4 months ago
llama3.2:latest   a806c4f17ac05 2.0 GB 4 months ago
llama3.1:latest   42182419e950 4.7 GB 4 months ago
> ollama run llama3.2
pulling manifest
Error: pull model manifest: file does not exist
> ollama run llama3.2
>>> what is the capital of Portugal
The capital of Portugal is Lisbon (Lisboa in Portuguese).
>>> [end a message (/? for help)]
```



# Knowledge Extraction from Databases

## Foundational Models (LLMs)



- **AI Models for Language**
  - LLMs (Large Language Models) are advanced AI models trained to understand and generate human-like text.
- **Massive Training Data**
  - They learn from vast amounts of text data, including books, articles, and code, to improve their responses.
- **Context Awareness**
  - LLMs understand context, allowing them to answer questions, summarize text, and generate creative content.
- **Continuous Learning**
  - Though they don't learn in real time, they can be fine-tuned or updated with new data to improve accuracy.
- **Wide Applications**
  - They power chatbots, search engines, code generation, content creation, and many other AI-driven tools.

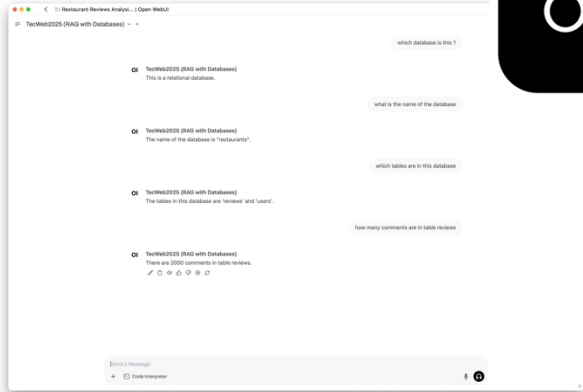
# Knowledge Extraction from Databases

## Architecture

Models (gpt4, etc...)



Models (llama3.2, etc...)

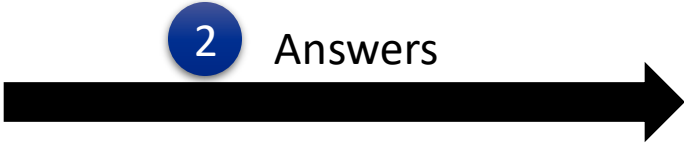


Graphical User Interface



Questions

1



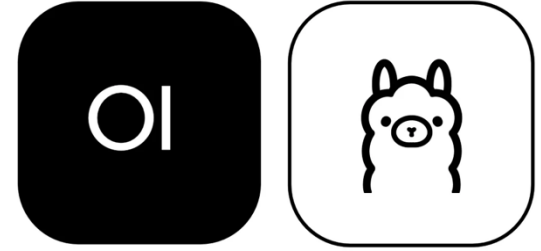
2

Answers



# Knowledge Extraction from Databases

## Graphical User Interface (OpenWeb UI)



- **Install & Set Up**
  - OpenWebUI is a frontend to connect to your preferred LLM (like OpenAI, LLaMA, or Mistral).
- **User-Friendly Chat Interface**
  - Provides a web-based chat interface where users can interact with LLMs through natural language.
- **Supports Multiple Models**
  - Works with various open-source and API-based LLMs, allowing easy switching between models.
- **Customizable & Extendable**
  - Allows users to modify UI settings, manage access, and integrate with other tools via APIs.
- **Self-Hosted & Private**
  - Runs on your own infrastructure, ensuring data privacy and control over AI interactions.

# Knowledge Extraction from Databases

## Databases



- **Relational Database**

- Is a structured way to store and manage data using tables.

- 1. Data is Stored in Tables**

- Information is organized into rows and columns, like a spreadsheet. Each table represents a specific entity (e.g., customers, orders).

- 2. Each Row is a Record**

- A row (or record) contains data for one instance of an entity (e.g., a single customer's details).

- 3. Each Column is a Field**

- Columns store specific types of information (e.g., "Name," "Email," "Phone Number").

- 4. Relationships Connect Tables**

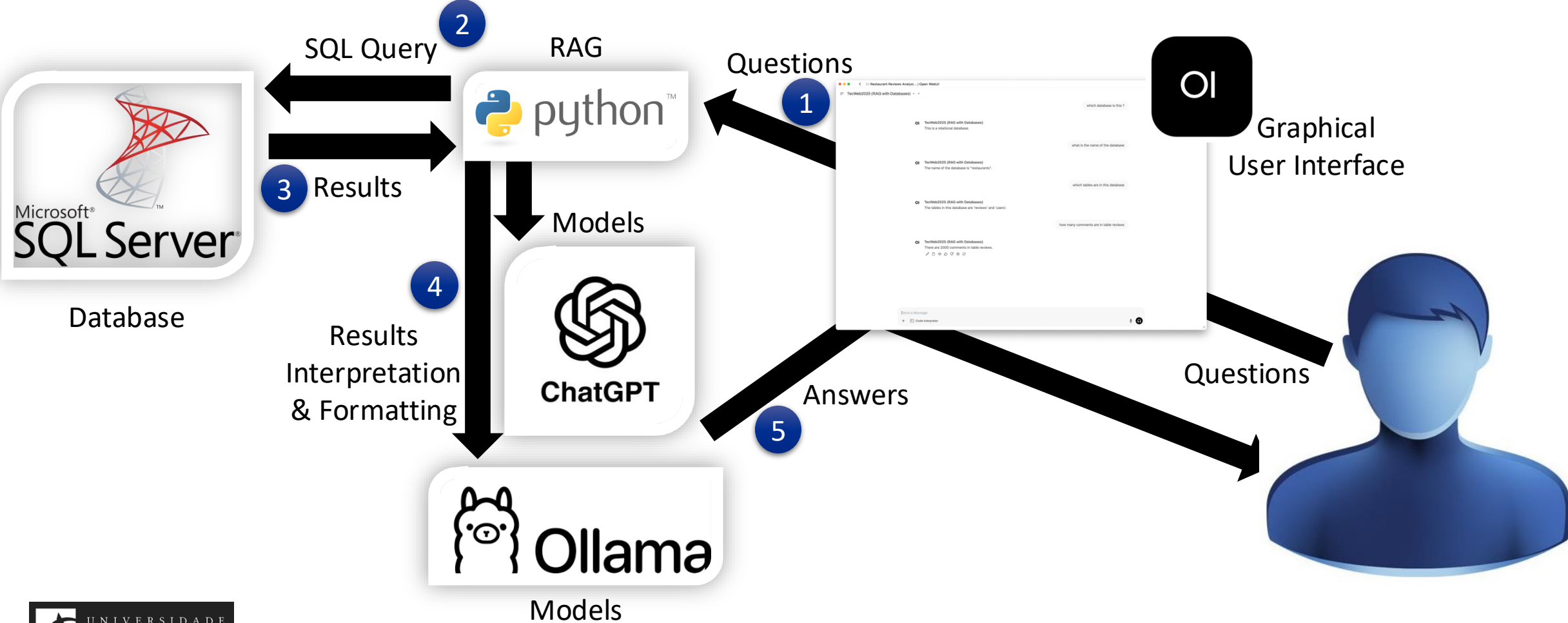
- Tables are linked using keys (e.g., a "Customer ID" in the orders table links to the customers table).

- 5. SQL is Used to Manage Data**

- Structured Query Language (SQL) is used to insert, update, delete, and retrieve data efficiently.

# Knowledge Extraction from Databases

## Architecture



# Knowledge Extraction from Databases

## Retrieval Augmented Generation (RAG)



- **Retrieval Augmented Generation (RAG)**

- Enhances language models by incorporating external knowledge retrieval.

- 1. User Query**

- The user provides a question or prompt to the system.

- 2. Retrieve Relevant Information**

- The system searches a knowledge base (e.g., documents, databases) to find relevant context.

- 3. Augment the Prompt**

- The retrieved information is added to the original query to provide more context.

- 4. Generate a Response**

- The language model (LLM) uses both the query and retrieved data to generate a more accurate answer.

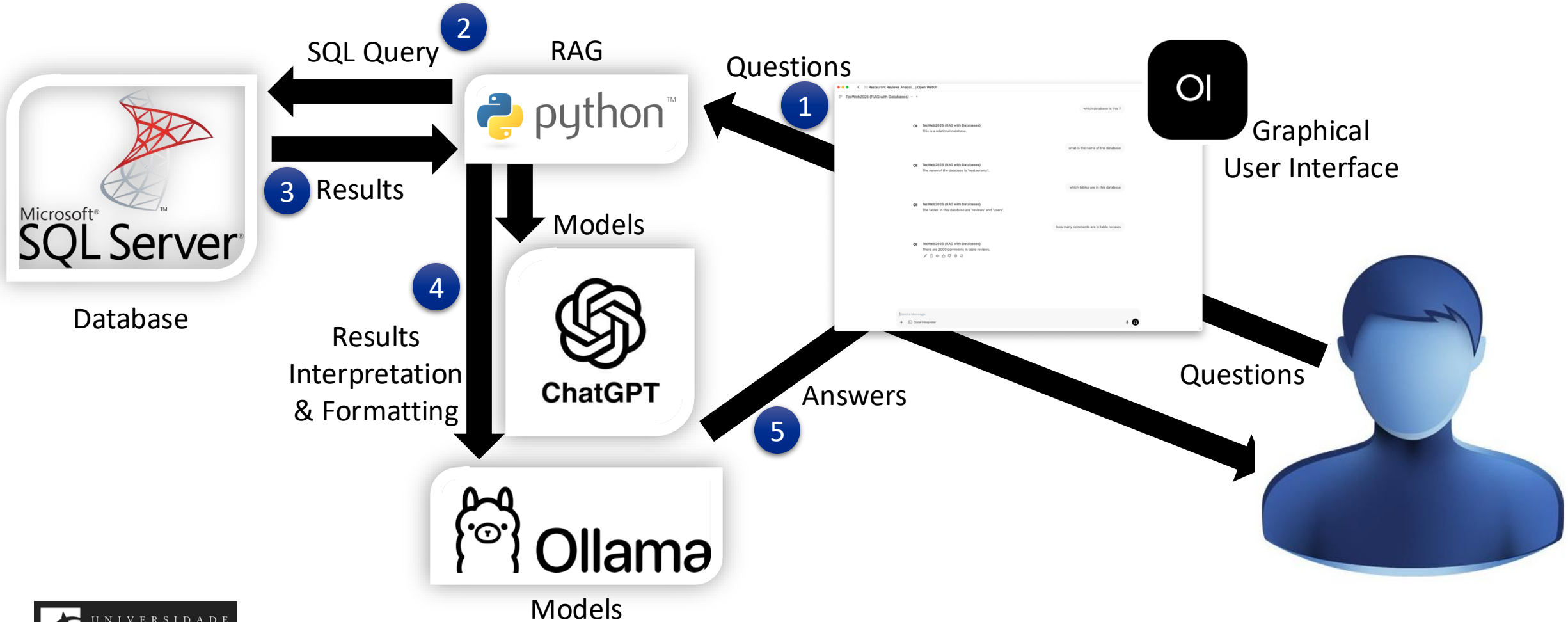
- 5. Deliver the Answer**

- The system returns a response that is more factual, relevant, and grounded in the retrieved data.



# Knowledge Extraction from Databases

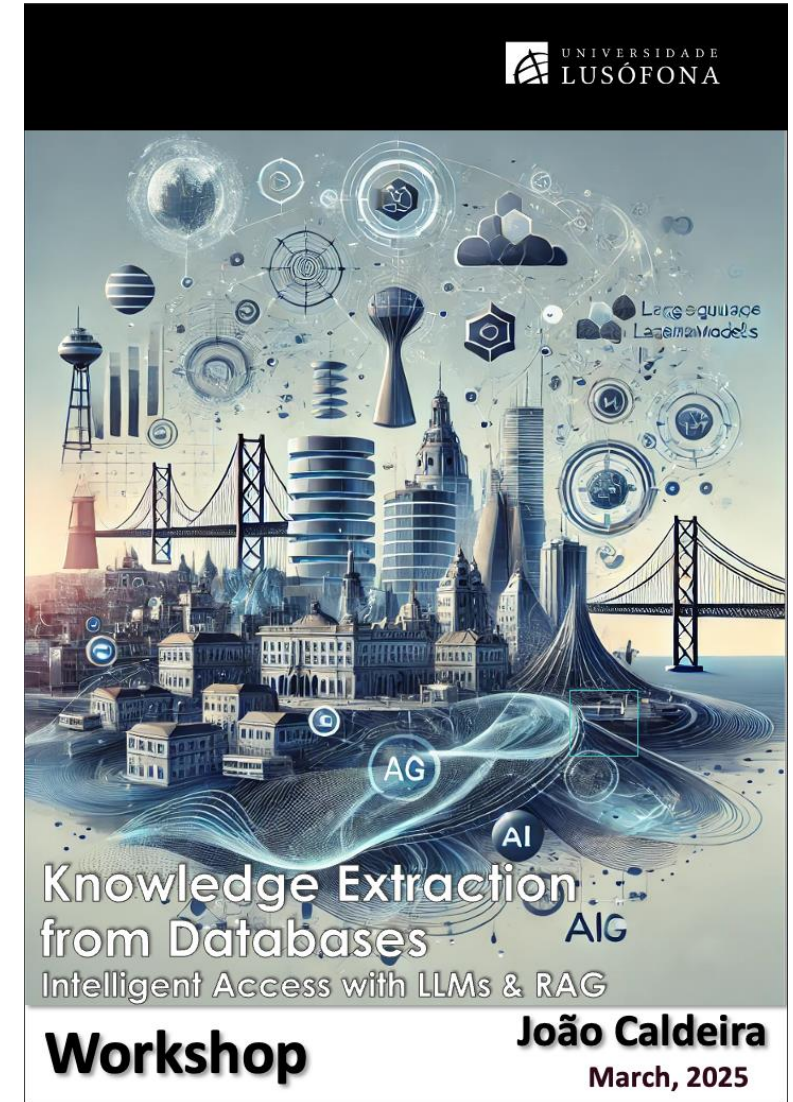
## Architecture



# Knowledge Extraction from Databases

## Documentation

- Access the documentation here:
  - <http://tecweb.kiion.com/2025>
  - Download in Pdf and Epub.



# Knowledge Extraction from Databases

**João Caldeira**

Assistant Professor

Email. [joao.caldeira@ulusofona.pt](mailto:joao.caldeira@ulusofona.pt)

March 19, 2025

