source allies

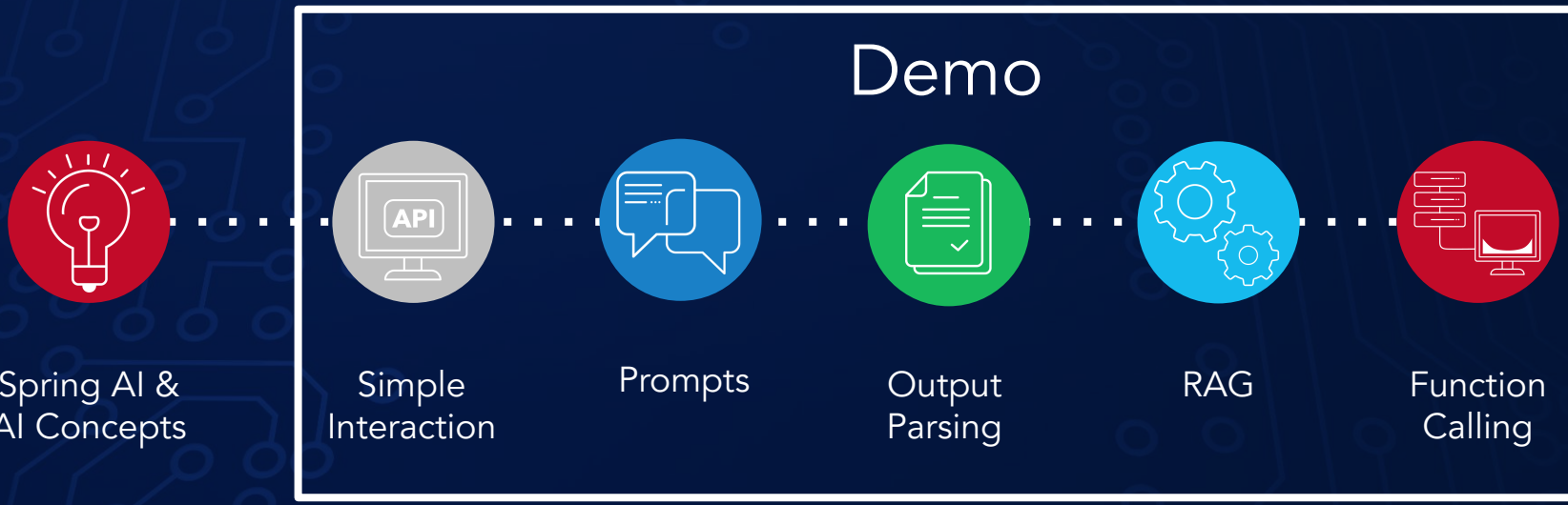
Leveraging LLMs with Spring AI

Spring AI

Kyle Hoehns

- Professionally tackling software problems since 2011
- Agriculture and Insurance Industries
- Source Allies since 2021

Agenda


Demo
API
Spring AI &
AI Concepts
Simple
Interaction
Prompts
Output
Parsing
RAG
Function
Calling

Spring AI

“The Spring AI project aims to streamline the development of applications that incorporate artificial intelligence functionality without unnecessary complexity.”

Timeline

- First commit in July 2023
- Experimental tags removed in December 2023
- Currently 1.0.0-M3

AI Models

Algorithms designed to process and generate information. Their training allows them to make predictions using insights from data sets.

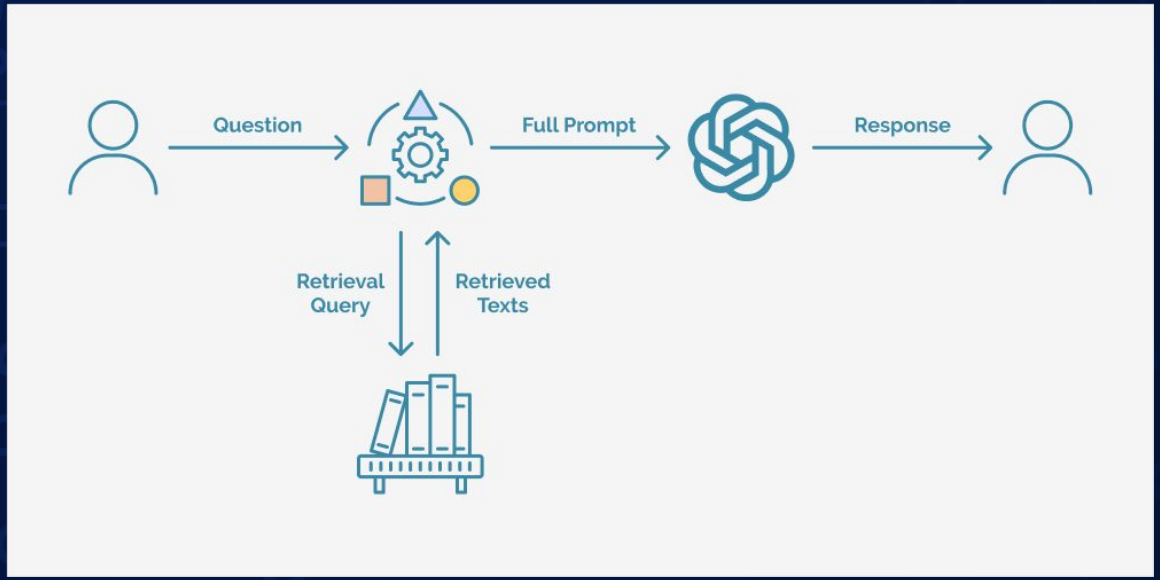
Spring AI Support:
Amazon Bedrock, Anthropic Claude, Azure Open AI, Google Vertex AI, Groq, HuggingFace, MistralAI, MiniMax, Moonshot AI, Ollama, OpenAI, QianFan, ZhiPu AI, and Watsonx.AI

Prompts

Building blocks for communication with AI Models that guide the model to producing relevant responses.

Retrieval-Augmented Generation (RAG)

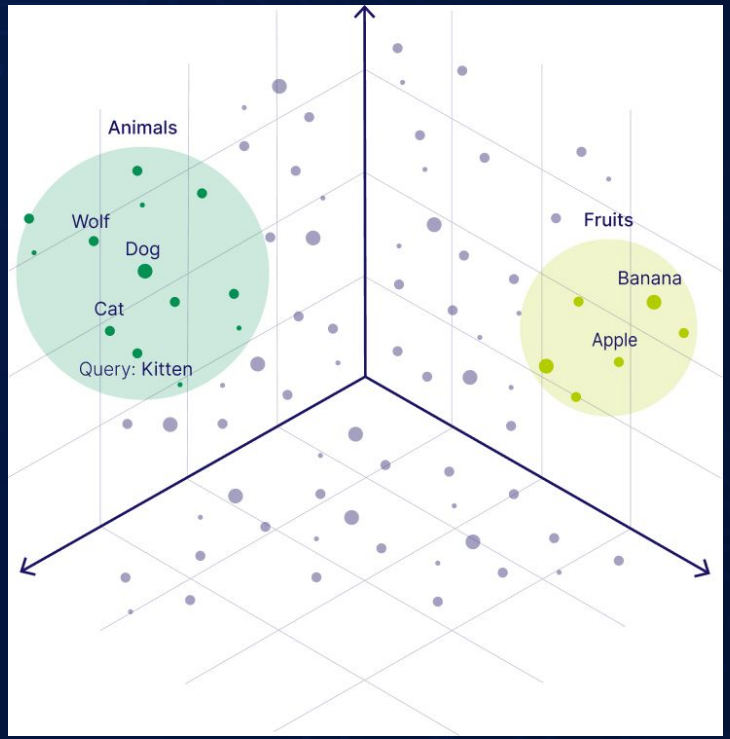
Store documents in vector storage related to data the model may not know about.



Vector Storage

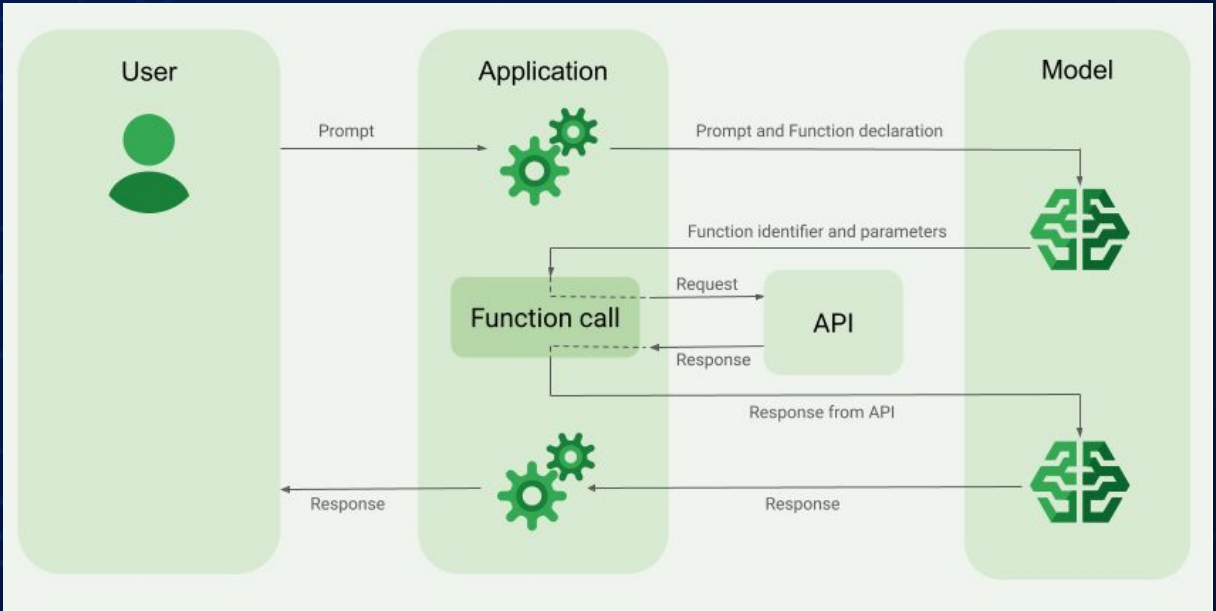
Store data alongside the data's vector embeddings.

Embeddings are arrays of floating point number that measure the relatedness of two entries.



Function Calling

Provide the model with the knowledge of other tools it can utilize to help answer questions.



Demo

Questions?

Resources

github.com/kylehoehns/spring-ai-demo

openai.com

docs.spring.io/spring-ai/reference