

Leveraging LLMs with Spring Al



Leveraging LLMs with Spring Al



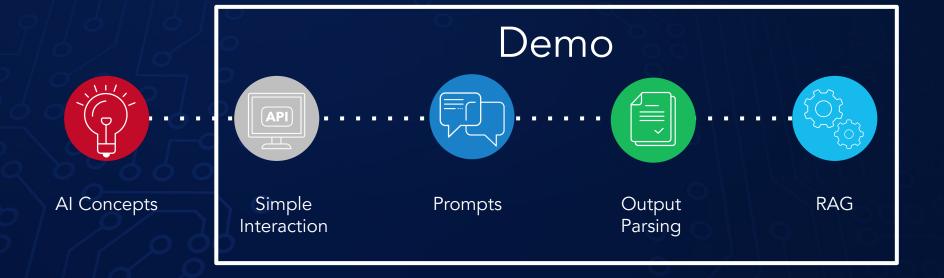
Kyle Hoehns

- Java/JVM Languages since 2011
- Agriculture and Insurance Industries
- Source Allies since 2021



Agenda





Concepts



Al Model

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

Current and Planned Spring AI Implementations

- Azure OpenAl
- Amazon Bedrock
- Hugging Face
- Ollama
- OpenAl
- PostgresML
- Vertex Al

Concepts



Prompts

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

Spring Al Assistance

Not all APIs for models require the same input.

Great templating support.





Retrieval-Augmented Generation (RAG)

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

Then retrieve and send that data to the model when asking questions.

Current Spring AI Vector Storage Implementations

- Azure
- Chroma
- Milvus
- Neo4j
- pgvector
- Pinecone
- Redis
- Weaviate

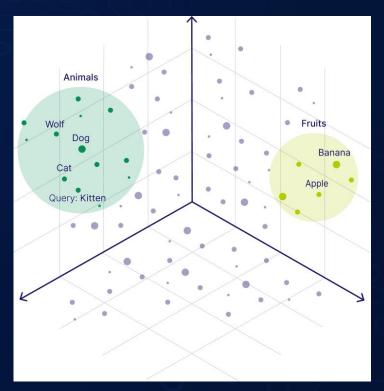
Concepts



Vector Databases

Store data alongside the data's vector embeddings.

Al Models help turn text/images/etc. into vector embeddings





Demo



Des Moines Artificial Intelligence Group

Des Moines Al Group meets the 3rd Tuesday of every month.

February 20th - Matt Vincent will be speaking on LLM Observability.







Questions?

Resources

github.com/kylehoehns/spring-ai-demo

openai.com

docs.spring.io/spring-ai/reference

