



source allies

Leveraging LLMs with Spring AI



Leveraging LLMs with Spring AI



Kyle Hoehns

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



Agenda



Demo



AI Concepts



Simple
Interaction



Prompts



Output
Parsing



RAG

Concepts



AI 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 OpenAI
- Amazon Bedrock
- Hugging Face
- Ollama
- OpenAI
- PostgresML
- Vertex AI

Concepts



Prompts

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

Spring AI Assistance

Not all APIs for models require the same input.

Great templating support.



Concepts

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.

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





Demo



Des Moines Artificial Intelligence Group

Des Moines AI 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

