

## **Codepocalypse Now:** LangChain4j vs. Spring AI



## **About the Speakers**



## **Baruch Sadogursky**

Head of Developer Relations at TuxCare.

- @jbaruch
- Co-author of "Liquid Software" and • "DevOps Tools for Java Developers"



## **Viktor Gamov**

Principal Developer Advocate at Confluent.

- @gamussa •
- Co-author of "Kafka in Action" •



### Java Expertise

Combined decades of Java development and advocacy experience.

- worldwide
- ecosystems

Speaking at major tech conferences

Leaders in AI integration with Java

## Shownotes

Access all presentation materials, code examples, and references at our resource hubs:



### **Baruch's Shownotes**

Comprehensive presentation materials and coding examples

speaking.jbaru.ch

Viktor's Shownotes

speaking.gamov.io

Same, just under different URL

### All You Need!

Includes slides, v code references

G.

@jbaruch

@gamussa

#J-Spring

speaking.gamov.io

speaking.jbaru.ch



#### Includes slides, video recording, and all



## **Vector Embeddings: The Semantic Backbone**

Words and phrases get mapped to high-dimensional vectors in mathematical space, capturing semantic relationships between concepts.



### **Vector Space Representation**

dimensional vectors in mathematical

Words and phrases get mapped to high-



## **Semantic Clustering**

Similar meanings cluster together in this vector space, allowing LLMs to understand relationships.



## **Semantic Relationships**

"automobile".



space.

*a*jbaruch

*a*gamussa *a* 

#J-Spring

speaking.gamov.io

speaking.jbaru.ch

### Vector spaces allow LLMs to understand that "happy" is closer to "joyful" than to



## Vector Math: The King-Queen Example

#### King - Man

Isolate the "royalty" concept by removing gender

### + Woman

Add the female gender component

## ≈ Queen

Result approximates the vector for "Queen"





@jbaruch

@gamussa

#J-Spring

speaking.ganpov.io





## **Conversational Memory in LLM Applications**

## **Context Window**

LLMs process a limited sequence of tokens as their "working memory".

## **Message Storage**

Applications store previous exchanges for continuity.

### **Memory Management**

Summarization techniques maintain essential information when context grows too large.

#J-Spring

speaking.gamov.io

speaking.jbaru.ch



## **Retrieval-Augmented Generation (RAG)**





*a*jbaruch

*a*gamussa

#J-Spring

speaking.gamov.io

speaking.jbaru.ch

## Query searches knowledge base for

System pulls relevant documents or





## **Docker Model Runner**

### Containerization

Packages LLMs with all dependencies in isolated containers.

Ensures consistent environment across deployments.

### Portability

## **Scalability**

Easily scale model instances to handle varying loads.

Orchestrate multiple models with container management tools.

#J-Spring

speaking.gamov.io

speaking.jbaru.ch

- Run the same model setup
- anywhere Docker is installed.
- Simplifies deployment from development to production.



## **MCP Servers: Model Context Protocol**









## **Thank You!**

- @jbaruch •
- @gamussa ٠
- #j-spring •
- speaking.jbaru.ch
- speaking.gamov.io





# **AI LANGUAGE MODELS AND CODING** </> </> </> </>