Lecture 01: Spring AI Introduction

What is Spring AI?

- Spring AI is a new application framework for AI engineering.
- It allows developers to integrate Generative AI features directly into Java/Spring Boot applications.
- Works as a layer between your enterprise application and LLMs (Large Language Models).
- Eliminates the need to switch between Java and Python for AI tasks.

Supported AI Providers:

- OpenAI (ChatGPT, GPT-4)
- Anthropic (Claude)
- Google (Gemini)
- Local models via Ollama etc.,

These providers expose Models as a Service (MaaS)—you connect using an API key and credits.

What Spring AI Provides?

AI allows you to integrate Generative AI into your app with ease. It supports:

- Chat Completion – build chatbots/Q&A systems
- Embeddings—convert text into vectors for semantic search
- Text → Image – generate images from prompts
- Audio Transcription – convert speech to text
- Text → Speech – generate speech from text
- Moderation – filter unsafe content

Key Takeaway:

- Spring AI is an AI engineering framework for Java.
- It makes adding AI-powered features like chat, search, image, and audio generation very simple if you already know Spring.