



Spring AI - Seamlessly Integrating AI into Your Enterprise Java Applications

Dr. Mark Pollack
Josh Long

Journey to Generative AI



Artificial Intelligence

Machine Learning

Deep Learning

Generative AI

1956 - Artificial Intelligence

CS field in developing intelligent machines that can think and learn like humans.

1997 - Machine Learning

Systems that learn from existing data to make decisions and predictions.

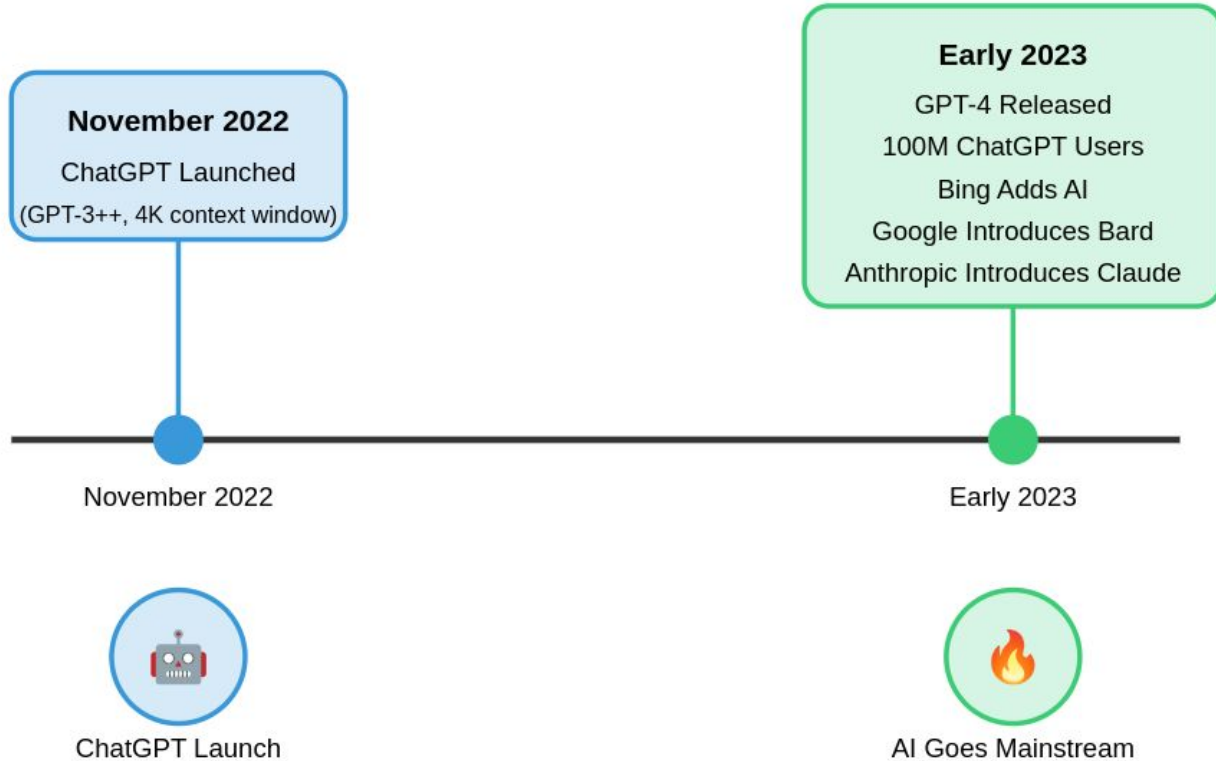
2012 - Deep Learning

Artificial Neural Networks, mimic the human brain structure, that can recognize complex patterns in data.

2021 - Generative Learning

Systems that can generate new data, such as images, videos, and audio.

Generative AI Timeline



Why Generative AI is Unique?

No training required

Human Language as the interface

No need to be a data scientist

Why can you do with it?

Have you had an 'ah-hah' moment?

Documents: Q&A, Summarization

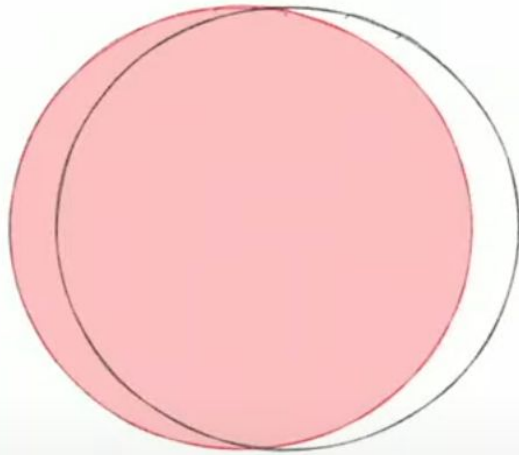
More intelligent chatbots

Fraud Detection

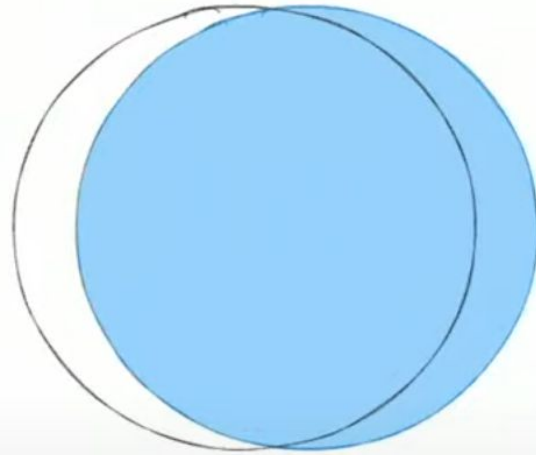
Developer Productivity

Evolution of the AI Engineer

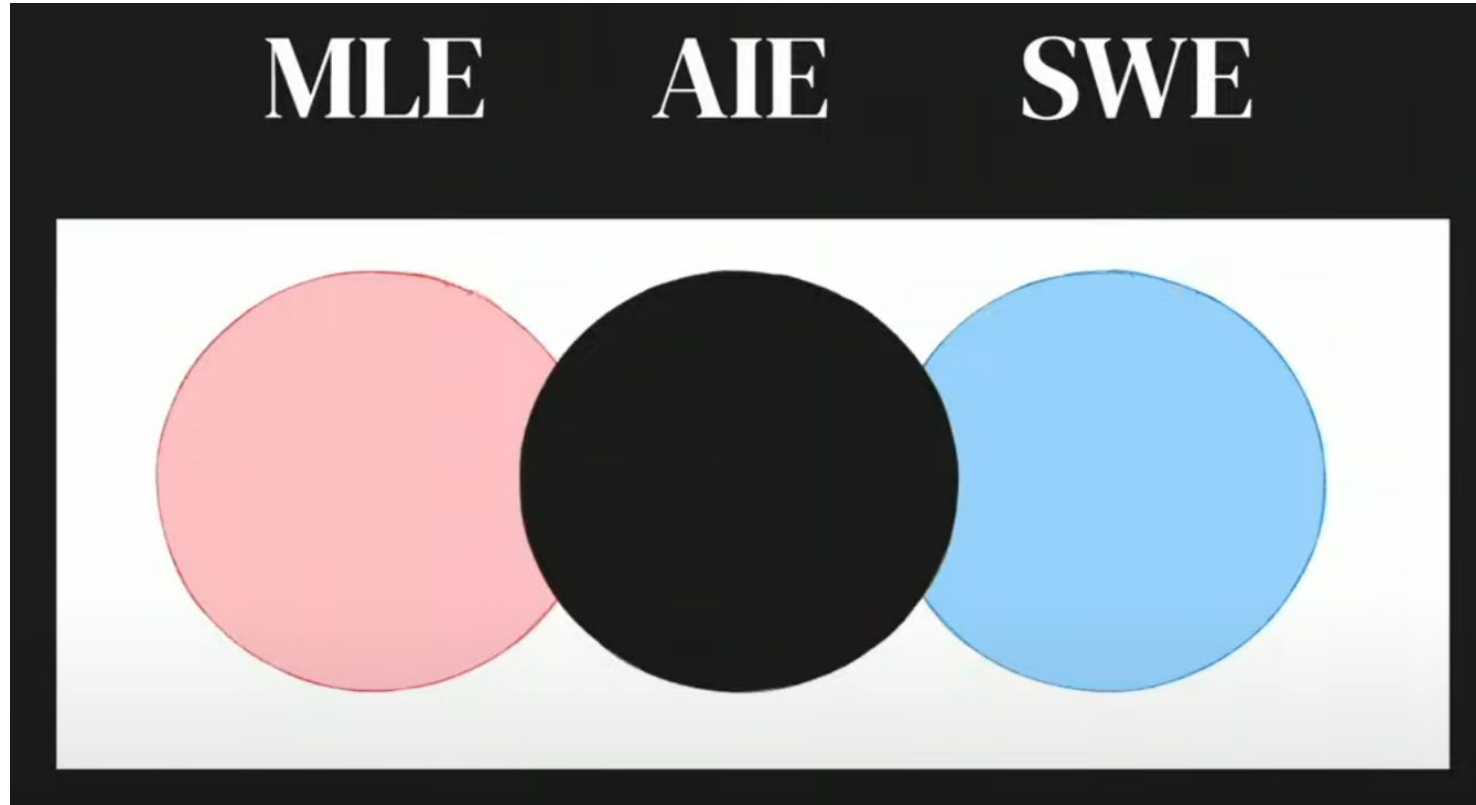
MLE POV



SWE POV



Evolution of the AI Engineer

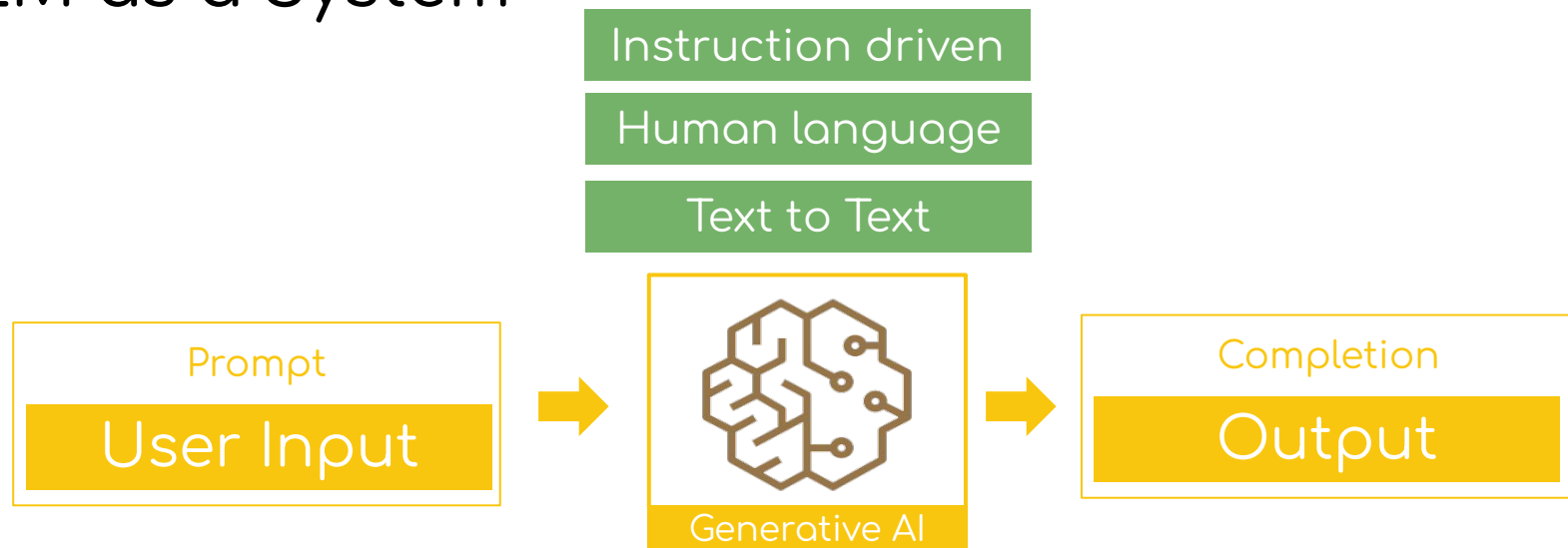


What is Spring AI?

An Application Framework for AI Engineering

- Portable AI Service Abstractions
- Developer Productivity
- Keeping up with the latest advancements

LLM as a System



Generative AI Patterns

Align responses to goals

Stateless APIs

Not aware of your APIs

Not trained on your data

Limited context size

Prone to hallucinations

System Prompt

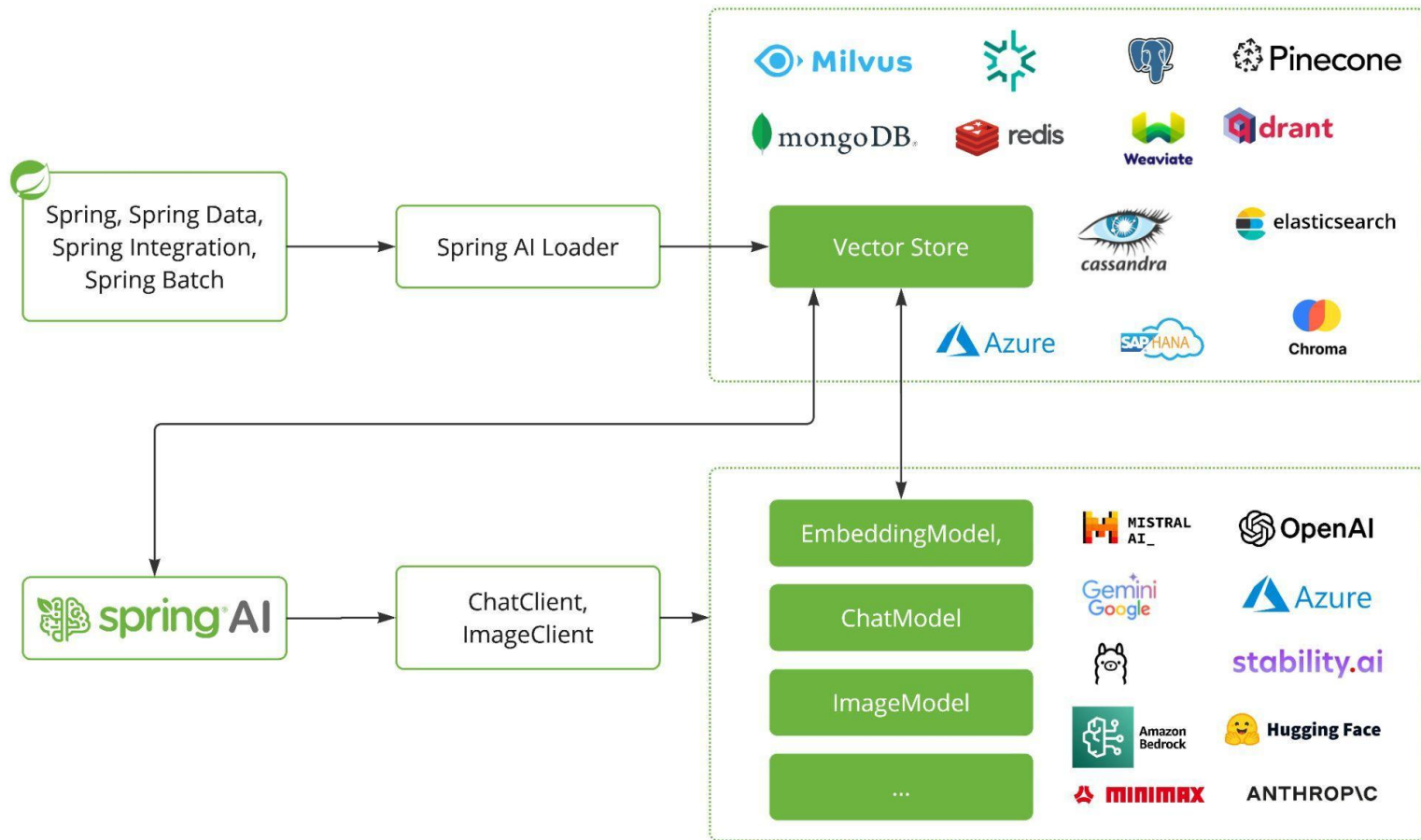
Chat Memory

Tool Calling

Prompt Stuffing

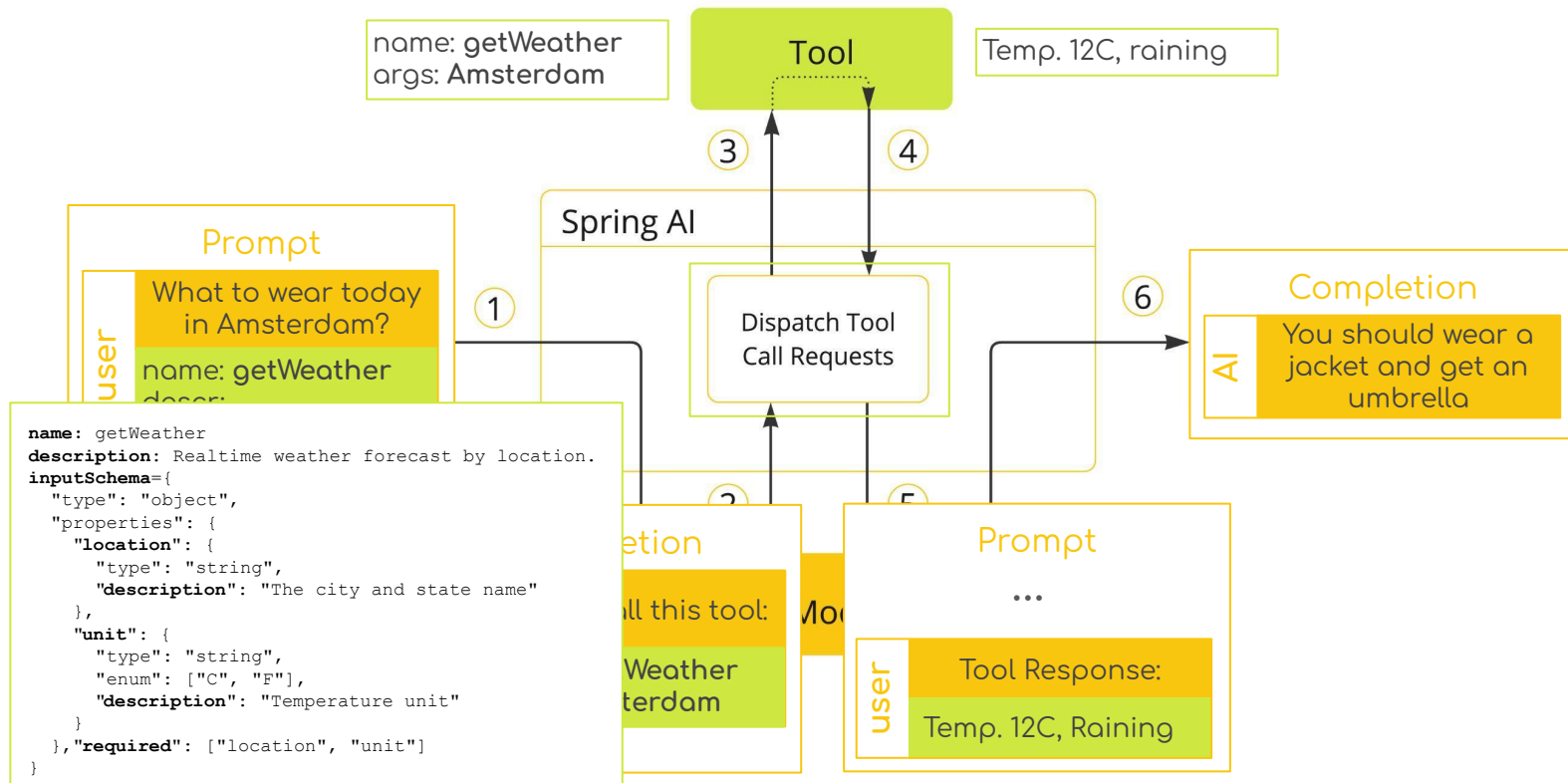
Vector Store + RAG

Evaluation

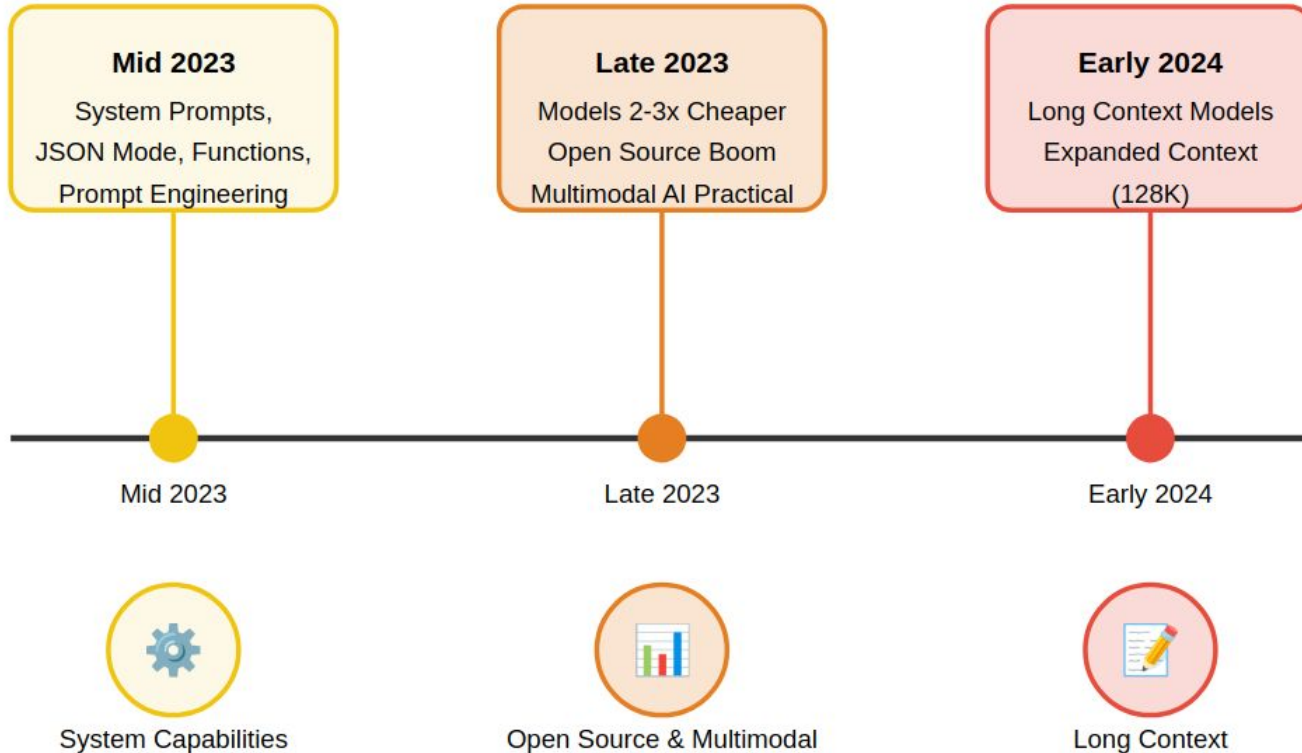


DEMO

Tool Calling



Generative AI Timeline



DEMO

Wouldn't it be nice...

- To be able to use tools written in other languages?
- To have Spring AI Tools used in non-Java envs?

Model Context Protocol (MCP)

Unified way to connect AI models to different data sources and tools

- Introduction



Search... 88K

Python SDK

TypeScript SDK

Java SDK

Kotlin SDK

Specification

Get Started

Introduction

Quickstart

Example Servers

Example Clients

Tutorials

Documentation

SDKs

Java SDK released! Check out [what else is new.](#)

MCP is an open protocol that standardizes how applications provide context to LLMs. Think of MCP like a USB-C port for AI applications. Just as USB-C provides a standardized way to connect your devices to various peripherals and accessories, MCP provides a standardized way to connect AI models to different data sources and tools.

Why MCP?

MCP helps you build agents and complex workflows on top of LLMs. LLMs frequently need to integrate with data and tools, and MCP provides:

- A growing list of pre-built integrations that your LLM can directly plug into
- The flexibility to switch between LLM providers and vendors
- Best practices for securing your data within your infrastructure

Local
Source A

Local
Source B

Internet

Remote
Service C

before.

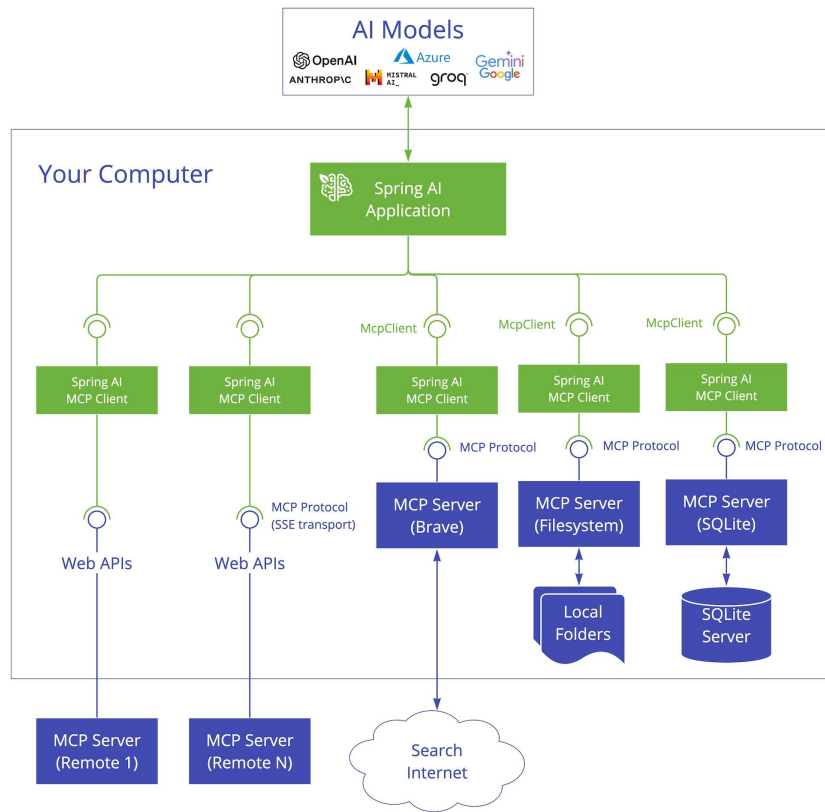


after.

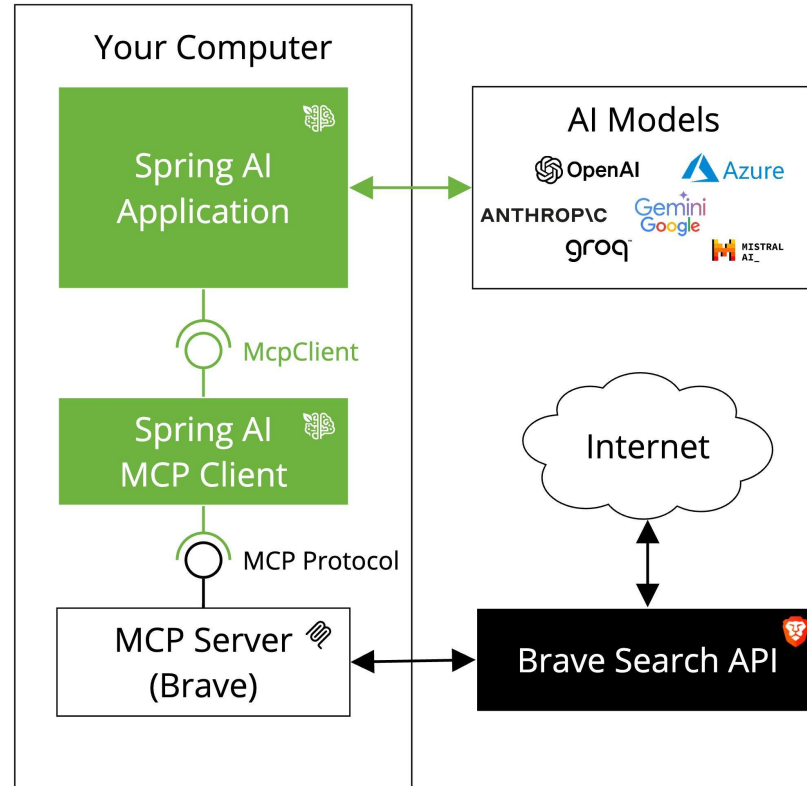


Java MCP Client

- Talks to Python, JS , .NET, Go, Java Servers
- Local vs Remote Servers
- Spring AI MCP Boot Starters



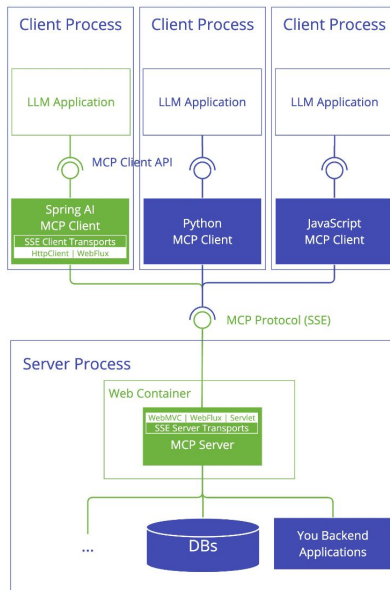
Spring AI + JavaScript WebSearch & Filesystem Tools



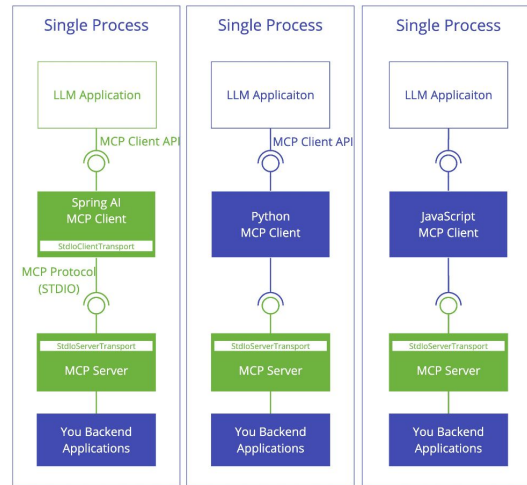
Java MCP Server

- Python/JS/Go.. MCP clients can connect to your Spring MCP Servers
- Spring AI MCP servers - collocated in user's applications (STDIO) or run remotely (SSE)

MCP Server (SSE)



MCP Server (STD IO)



A new toolbox...

The screenshot shows a web browser window with the URL `www.mcpserver.info`. The page has a dark theme and features a prominent title "Model Context Protocol Server Directory" in white. Below the title, it states "Curated directory of trusted MCP servers (10+ ★)" and provides a brief description of the directory's purpose. A search bar and a filter button labeled "All" are visible. A grid of category buttons includes "MCP Tools", "Development", "System Operation", "Web", "Cloud", "Data", "Knowledge Base", "Communication", "Content Tools", "Business", "paraDX", and "The Others". The bottom section displays three featured server cards: "jlowin/fastmcp" (highlighted as No. 1 >1000), "chrisayuk/mcp-cli", and "cloudflare/mcp-server-cloudflare". Each card includes a profile picture, repository statistics, a description, a category tag, and buttons for "View Details" and "View Repo".

Private < >

www.mcpserver.info

Subscribe

Model Context Protocol Server Directory

Curated directory of trusted MCP servers (10+ ★)

Navigate the rapidly growing Model Context Protocol ecosystem with confidence. From 600+ servers launched in the first month, we highlight the most established and community-verified ones.

Search...

All About MCP

MCP Tools Development System Operation Web Cloud Data

Knowledge Base Communication Content Tools Business paraDX The Others

jlowin/fastmcp

The fast, Pythonic way to build Model Context Protocol servers

6 Contributors 126 Used by 2 Discussions 1k Stars 54 Forks

🏆 FastMCP! No. 1 >1000 ★

The fast, Pythonic way to build Model Context Protocol servers

Development

View Details View Repo

chrisayuk/mcp-cli

8 Contributors 10 Issues 632 Stars 126 Forks

🏆 MCP CLI

Command Line Interface for interacting with Model Context Protocol servers

MCP Tools

View Details View Repo

cloudflare/mcp-server-cloudflare

4 Contributors 1 Used by 329 Stars 25 Forks

MCP Server Cloudflare

The bridges between LLMs and Cloudflare's API infrastructure

Cloud

View Details View Repo

DEMO

Agents...



Agentic AI System

“Leverages an **AI model** to **interact with** its **environment** in order to **solve** a **user-defined task**”

Combines **Planning** & **Memory** & **Actions** to fulfill the tasks



AI Model

Handles reasoning and planning
Decides which Actions to take



Tools

What an Agent is equipped to do
Interacts with its environment

Agentic Systems Types

- Workflows - LLMs and Tools are orchestrated through predefined paths



Fully Autonomous Agents are appealing,

OUT

- Agents - LLMs and Tools are orchestrated through predefined paths and e.g. autonomous agents towards accomplishing the tasks - e.g. autonomous

Workflows offer better Predictability and Consistency for well-defined tasks.

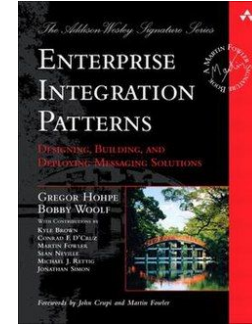
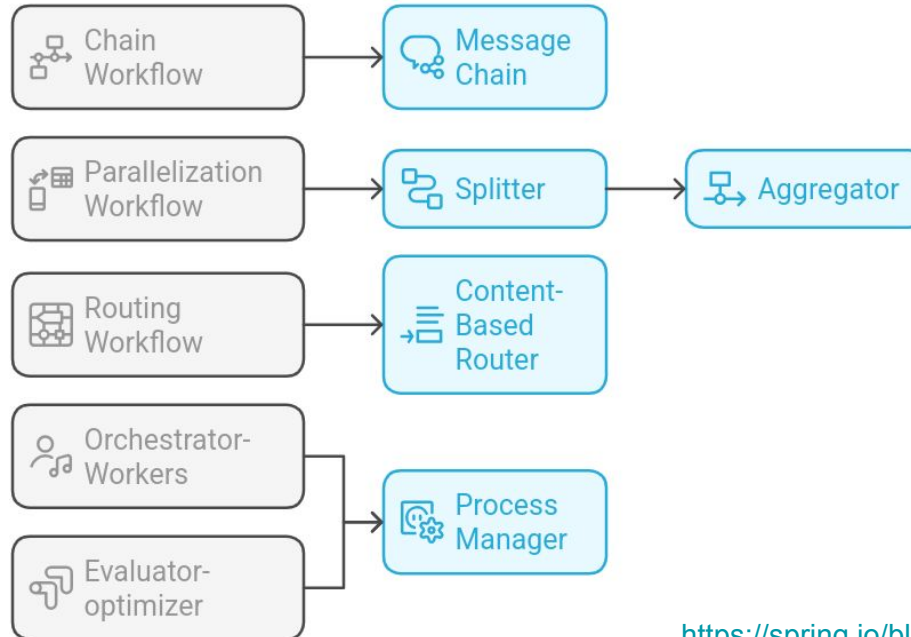
Retrieval

Tools

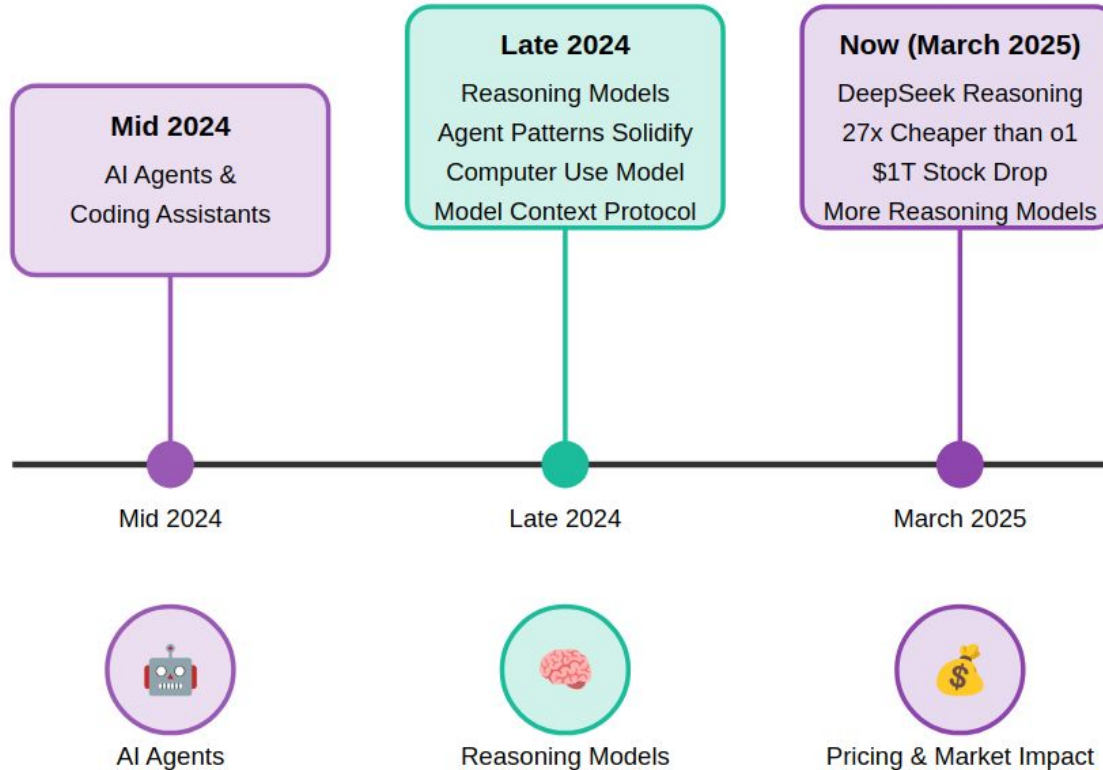
Memory

Agent workflow patterns

- Similar to Enterprise Integration Patterns...



Generative AI Timeline



Thank You

Resources

Docs

- <https://docs.spring.io/spring-ai/reference/index.html>
- <https://docs.spring.io/spring-ai-mcp/reference/overview.html>
- <https://spring.io/blog/2025/01/21/spring-ai-agentic-patterns>

Demos

- <https://github.com/spring-projects/spring-ai-examples>
- <https://github.com/joshlong-attic/2025-devnexus-bootiful-ai>