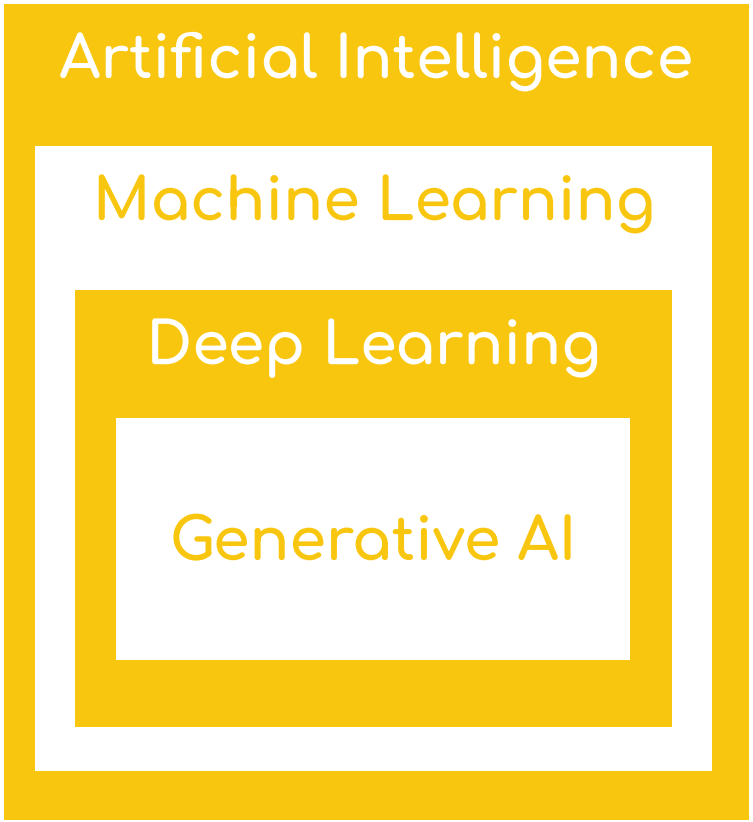
Spring AI - Seamlessly Integrating AI into Your Enterprise Java Applications

Dr. Mark Pollack
Josh Long

Journey to 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**.

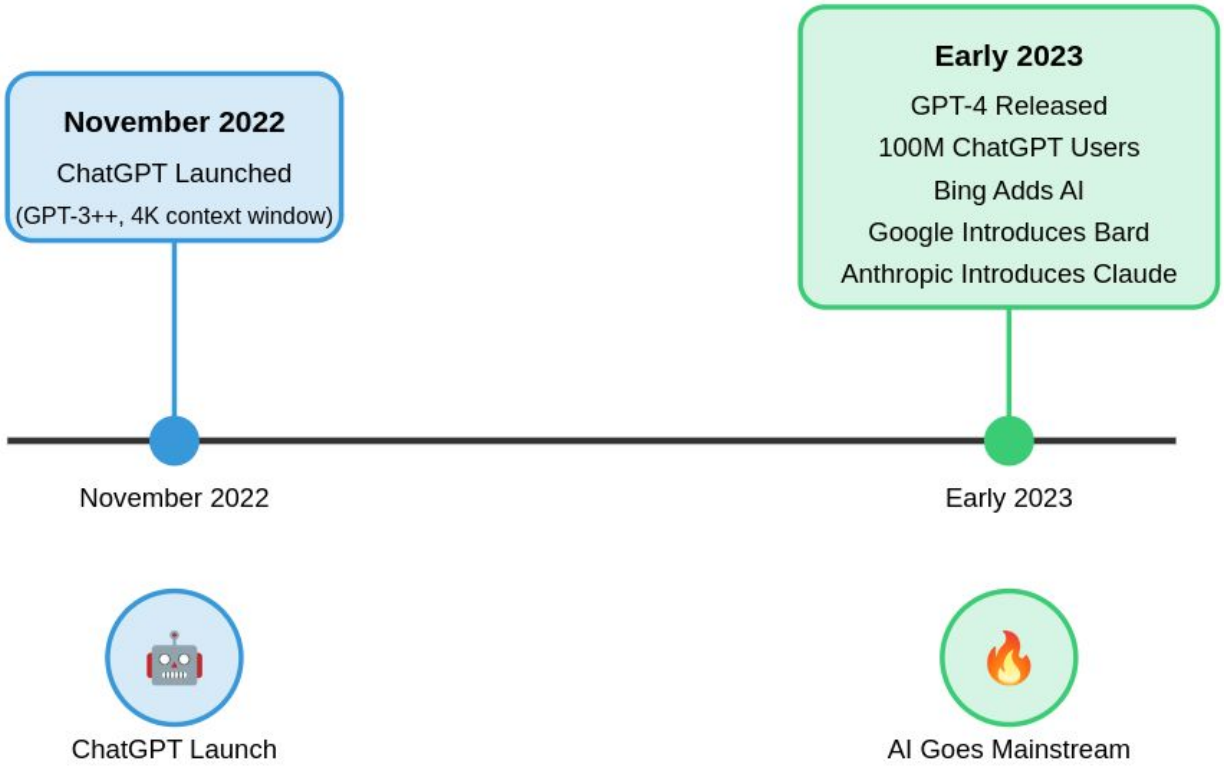
2012 - Deep Learning

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?

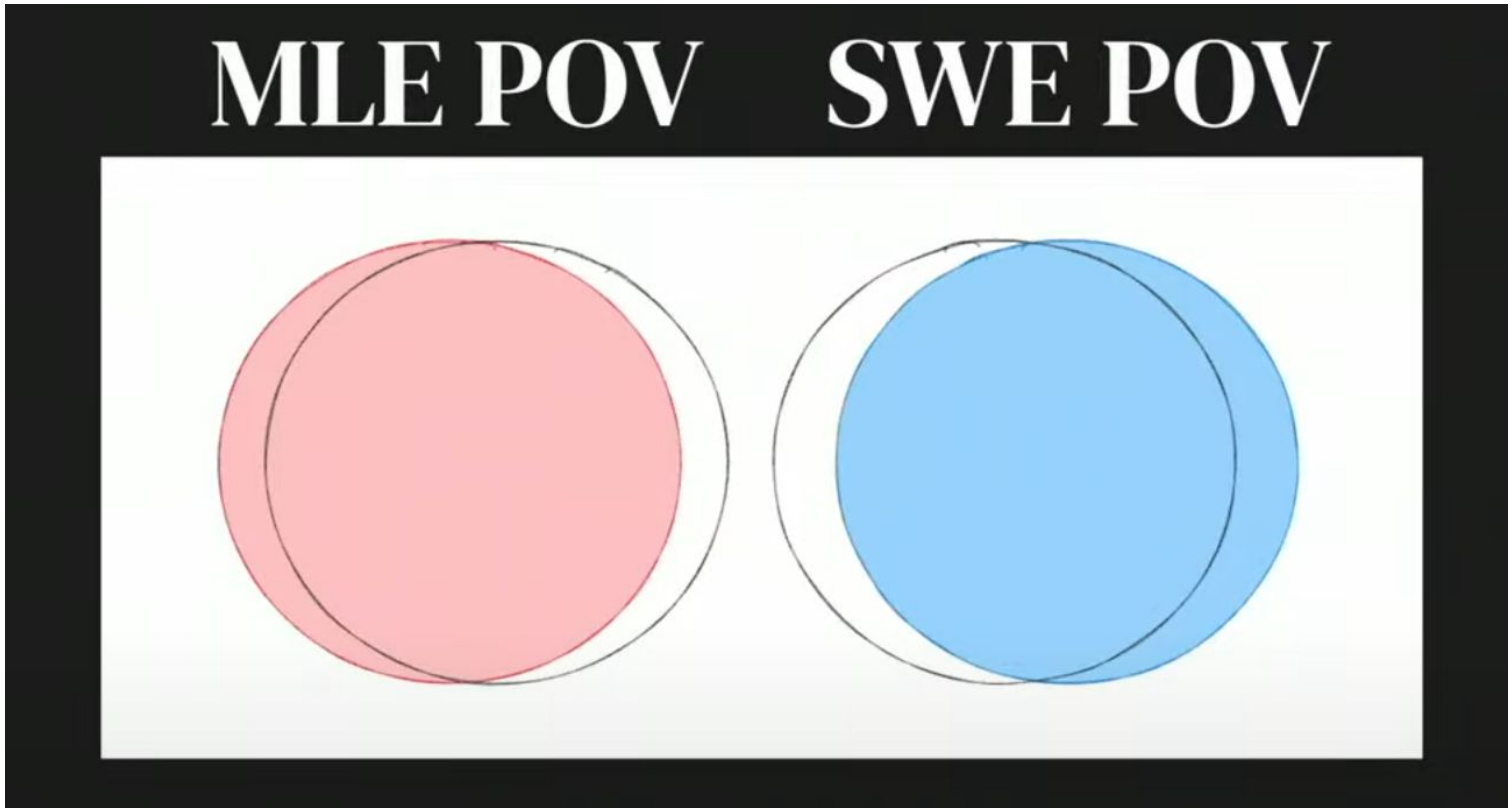
Documents: Q&A, Summarization

More intelligent chatbots

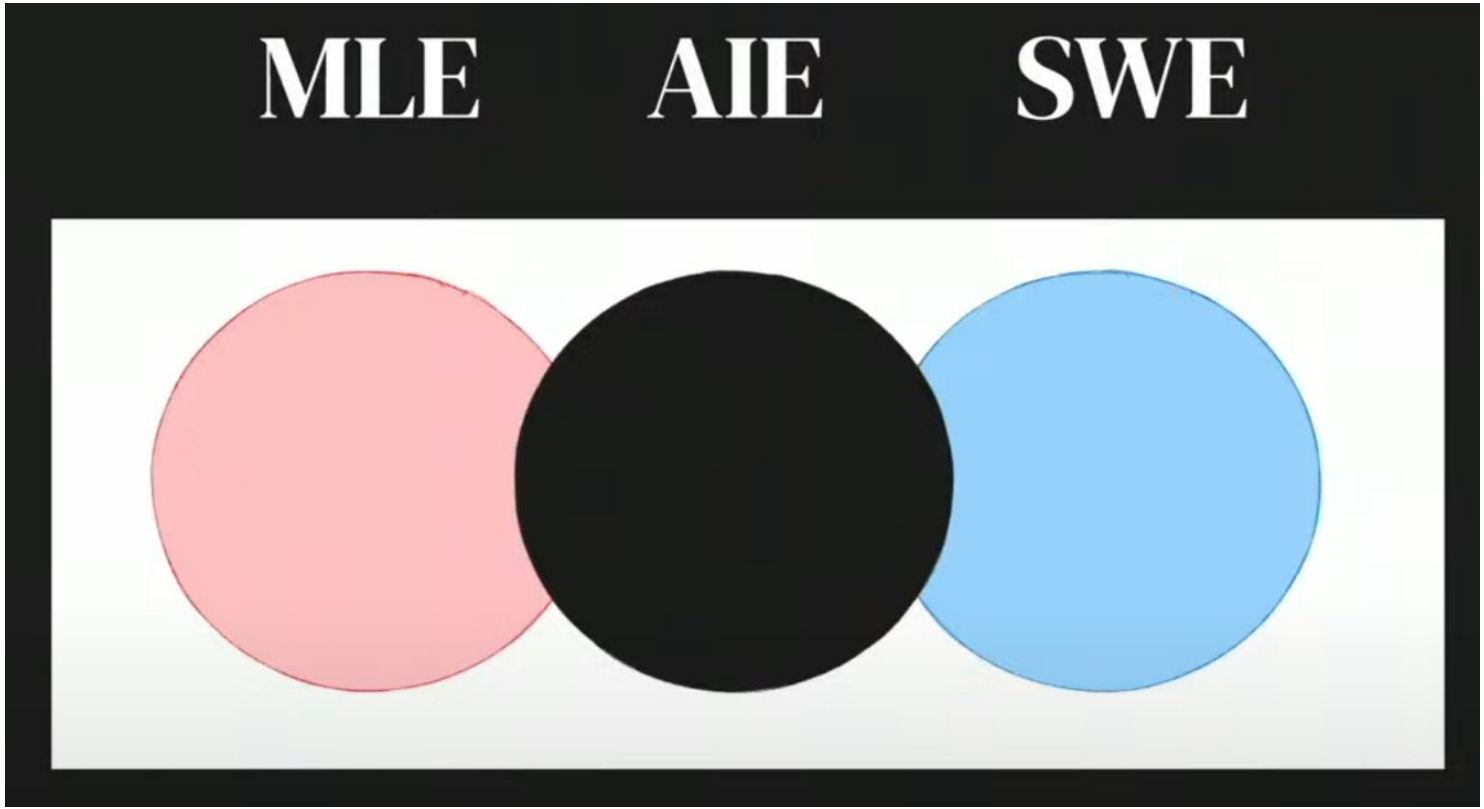
Fraud Detection

Developer Productivity

Evolution of the AI Engineer



Evolution of the AI Engineer

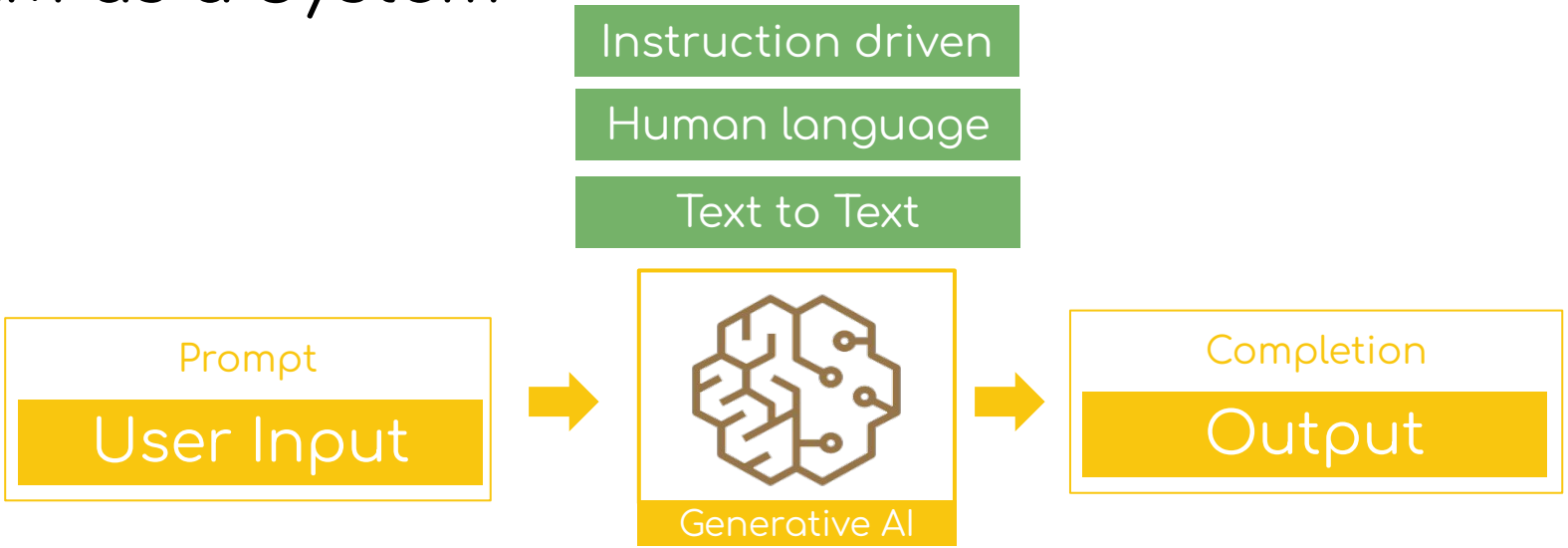


What is Spring AI?

An Application Framework for AI Engineering

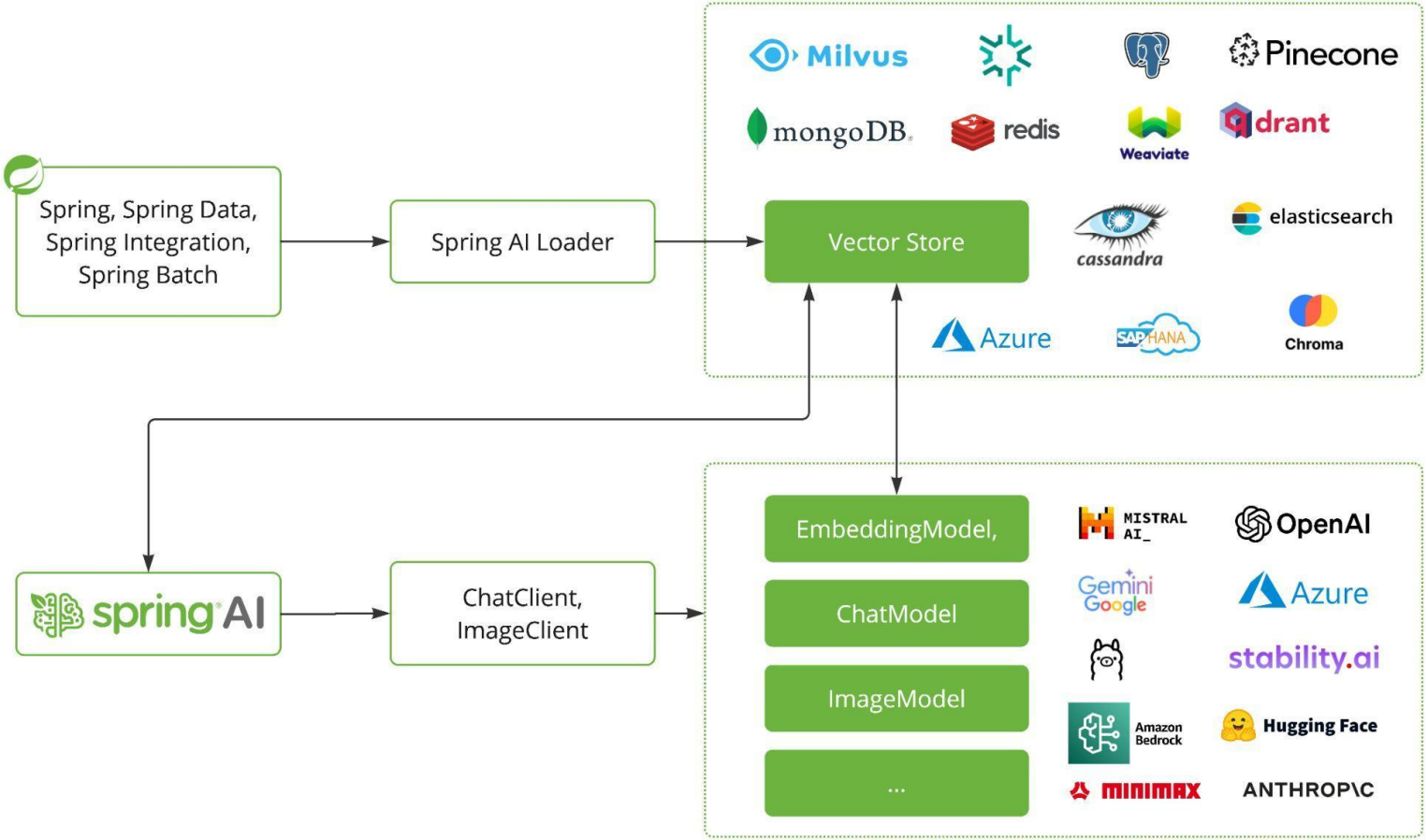
- Portable AI Service Abstractions
- Developer Productivity
- Keeping up with the latest advancements

LLM as a System



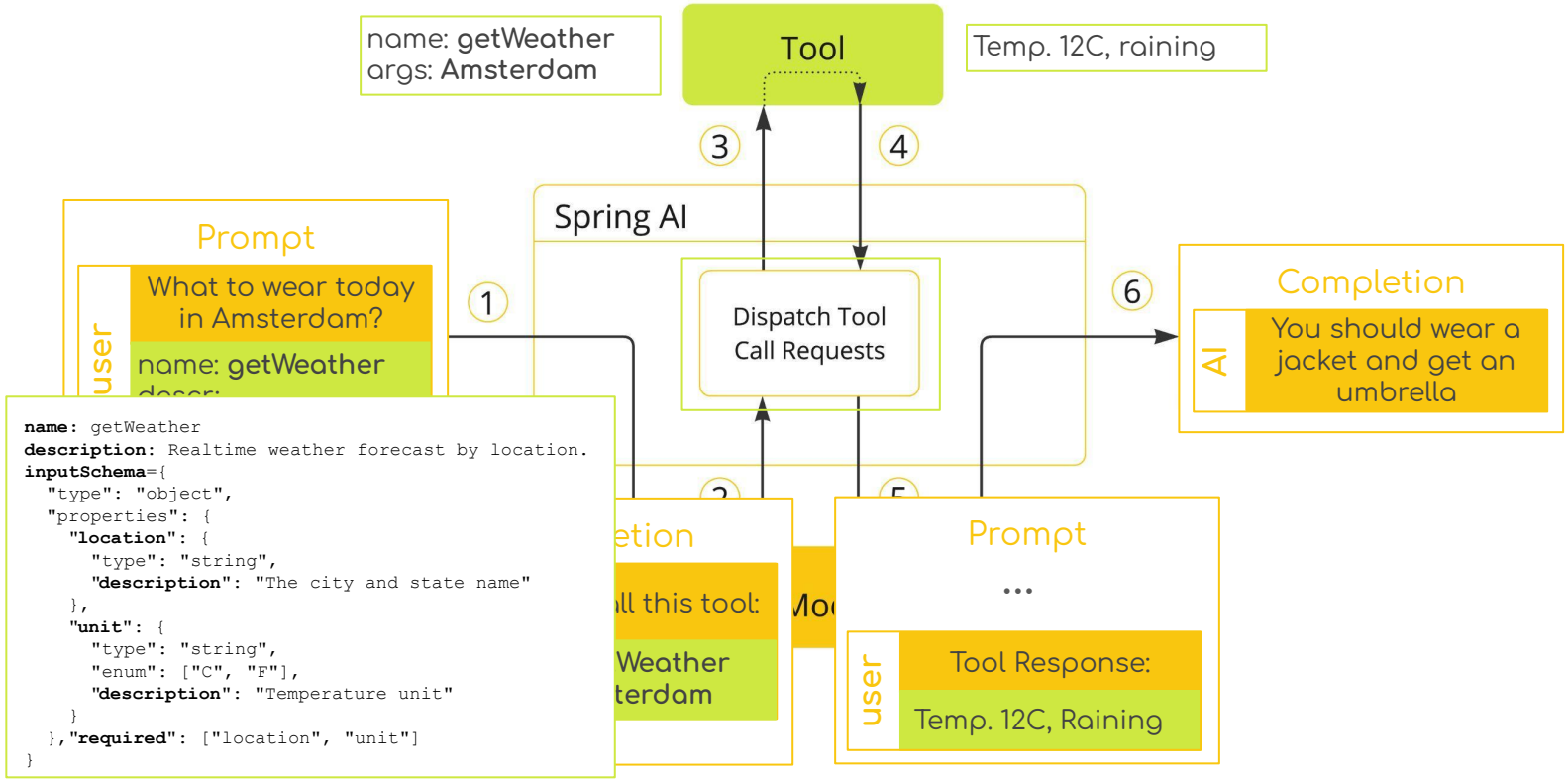
Generative AI Patterns

Align responses to goals	System Prompt
Stateless APIs	Chat Memory
Not aware of your APIs	Tool Calling
Not trained on your data	Prompt Stuffing
Limited context size	Vector Store + RAG
Prone to hallucinations	Evaluation

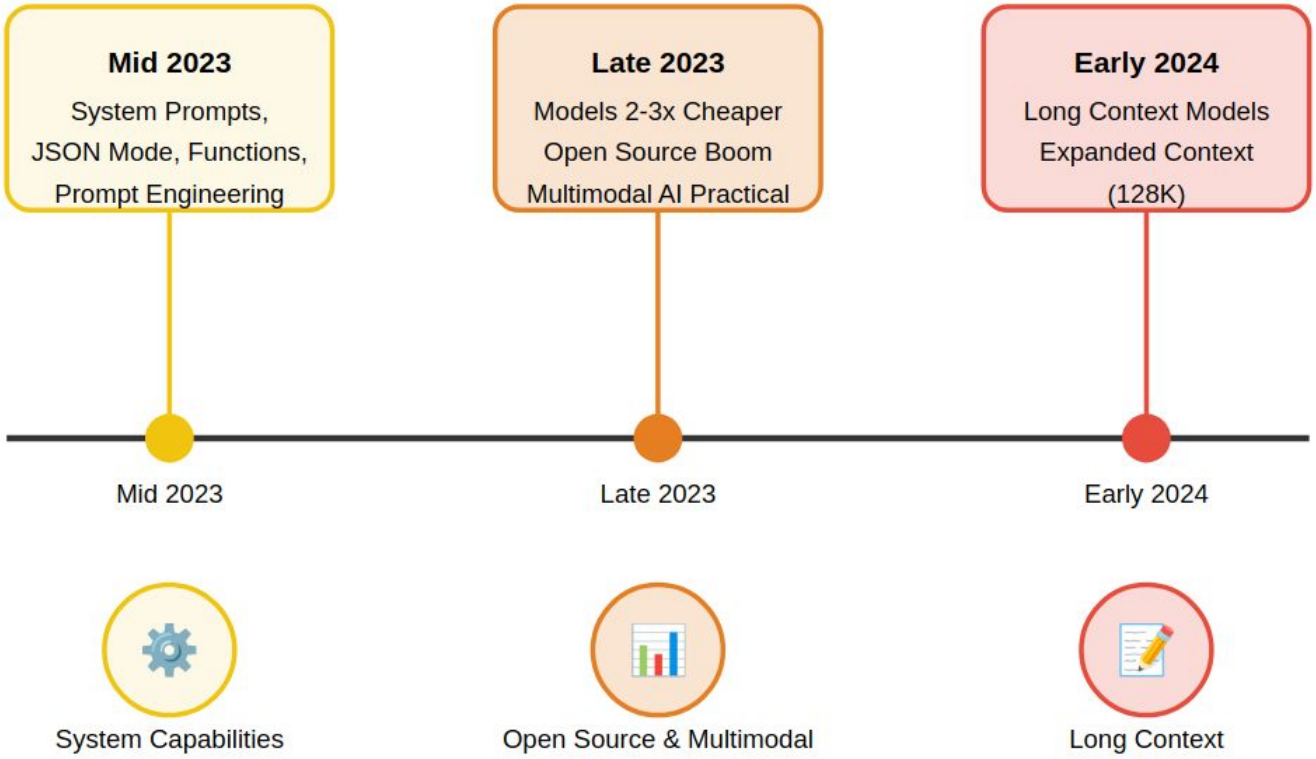

Spring, Spring Data, Spring Integration, Spring Batch
Spring AI Loader
Vector Store
Milvus
Pinecone
mongoDB.
redis
Weaviate
drant
cassandra
elasticsearch
Azure
SAP HANA
Chroma
spring AI
ChatClient, ImageClient
EmbeddingModel,
ChatModel
ImageModel
...
MISTRAL AI_
OpenAI
Gemini Google
Azure
stability.ai
Amazon Bedrock
Hugging Face
MINIMAX
ANTHROP\C

DEMO

Tool Calling



Generative AI Timeline



DEMO

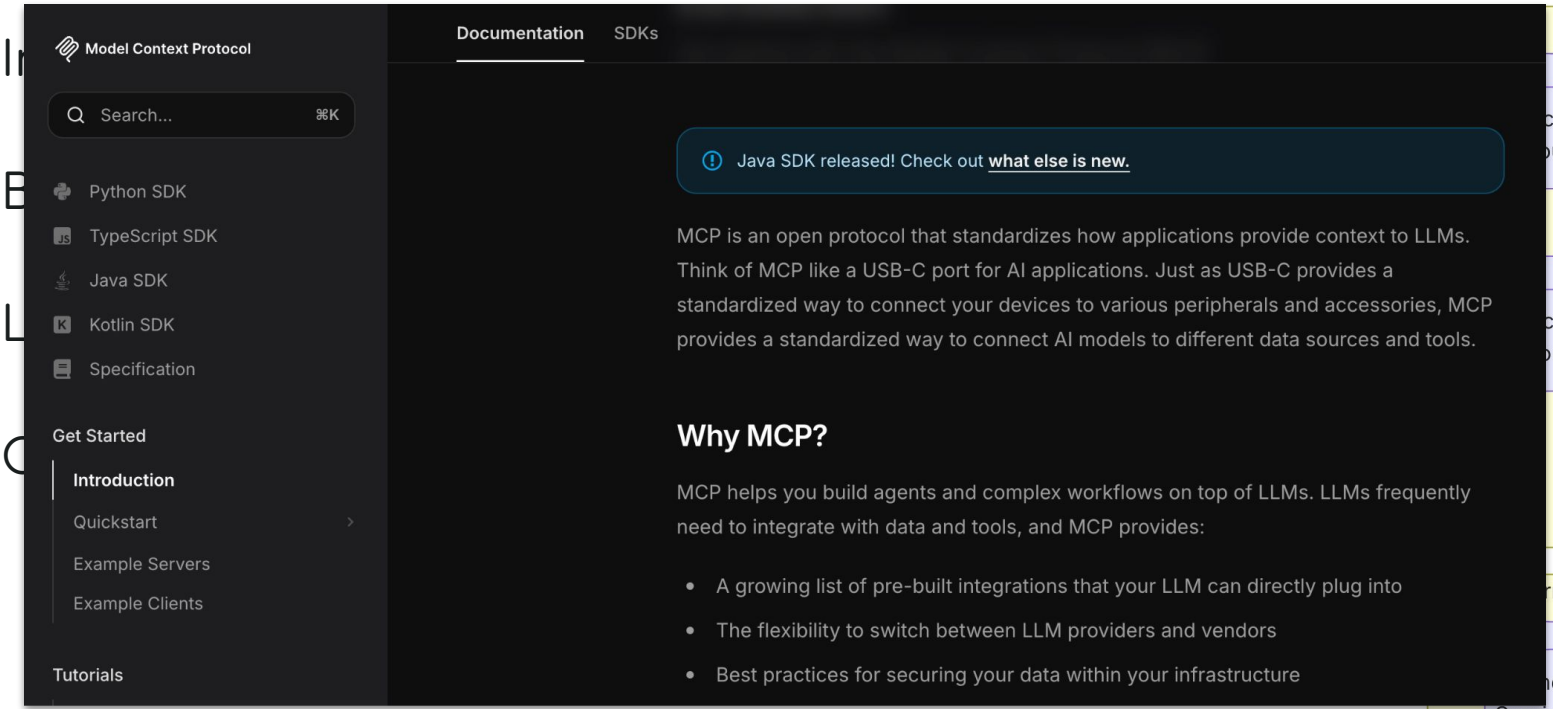
Wouldn't it be nice...

- 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**

- I
- B
- L
- C
  - ...

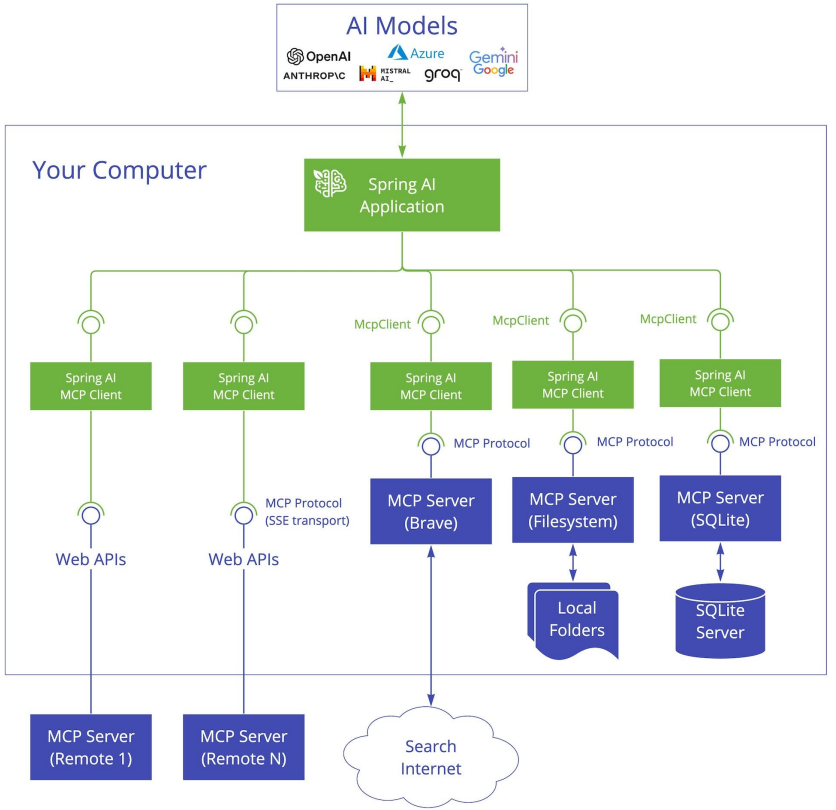


before.

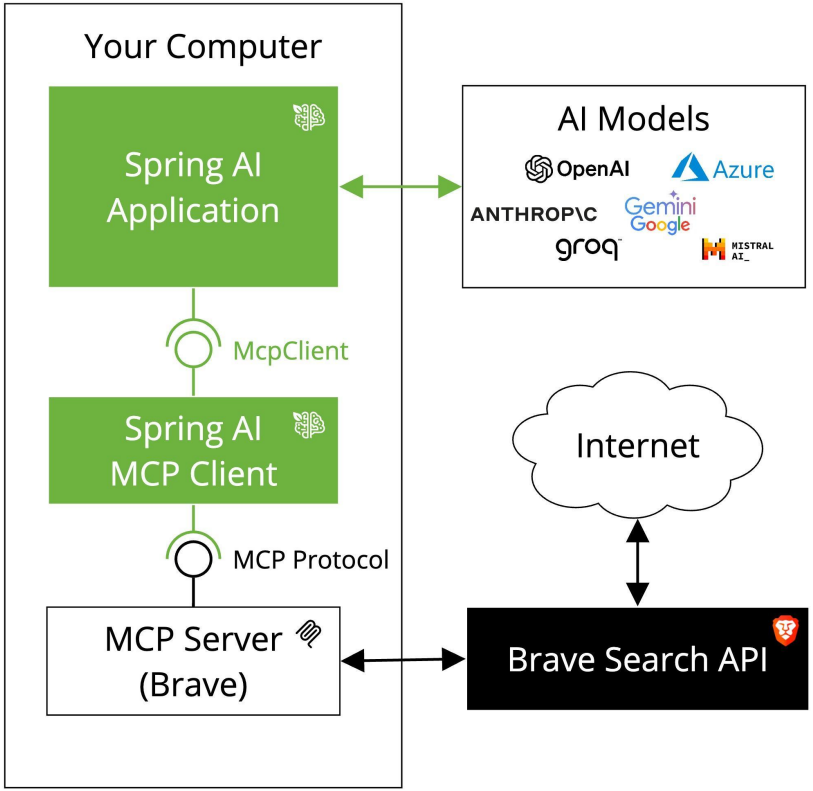
after.

Java MCP Client

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



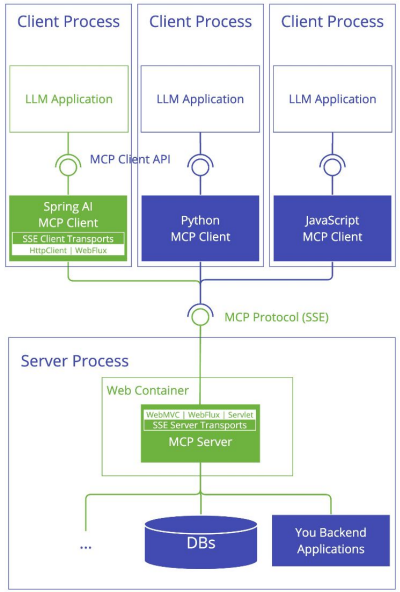
Spring AI + JavaScript WebSearch & Filesystem Tools



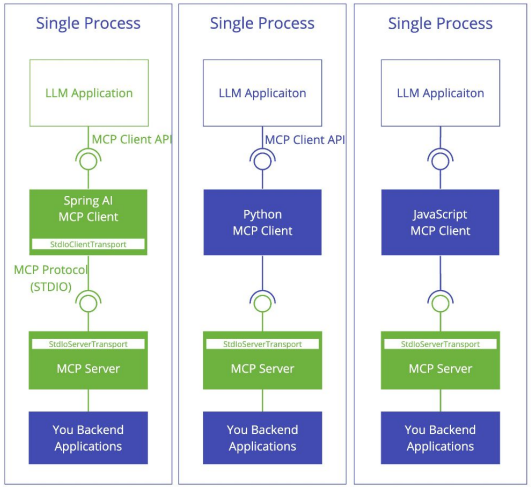
Java MCP Server

- 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)

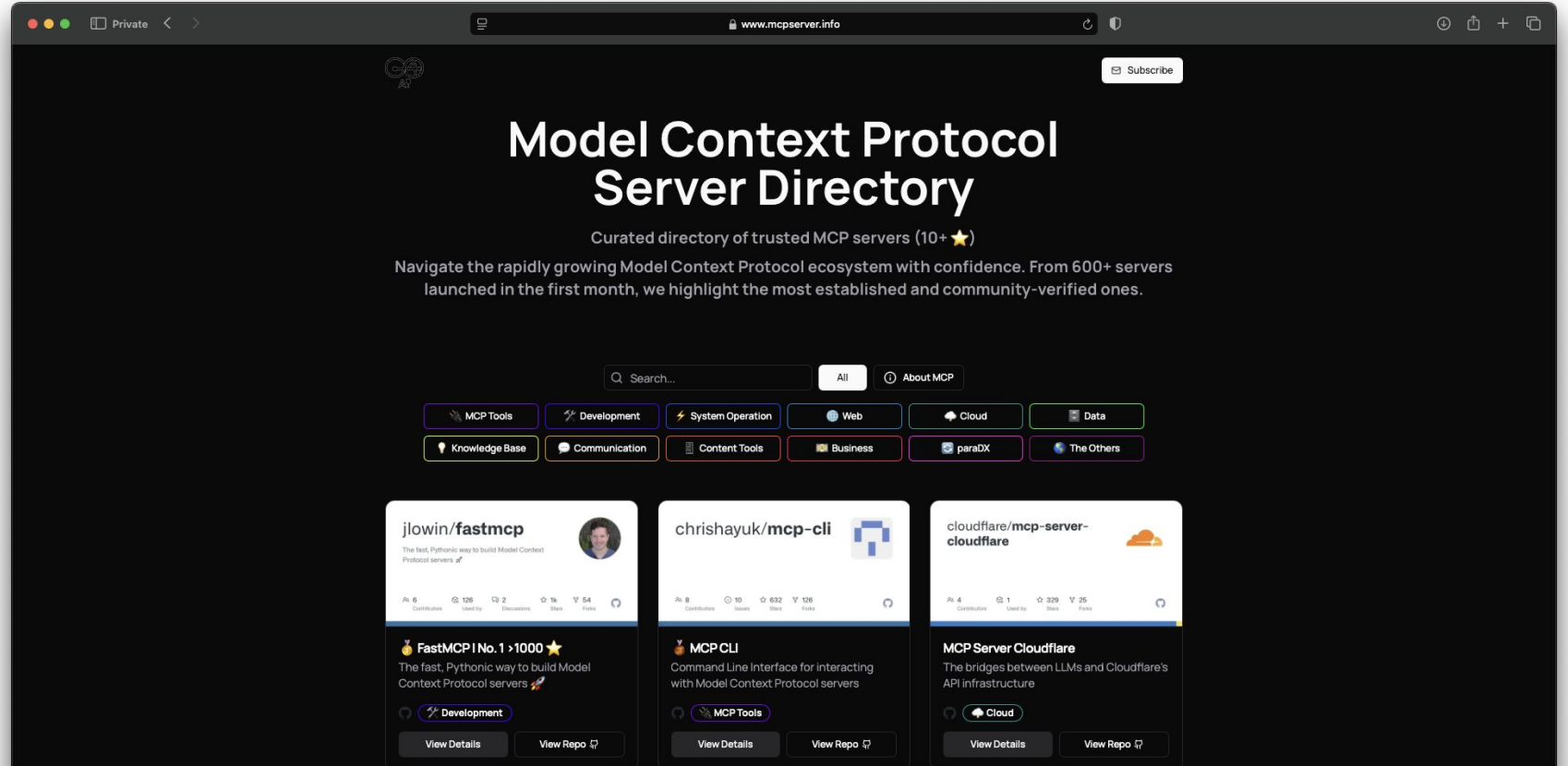
MCP Server (SSE)



MCP Server (STD IO)



A new toolbox...



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

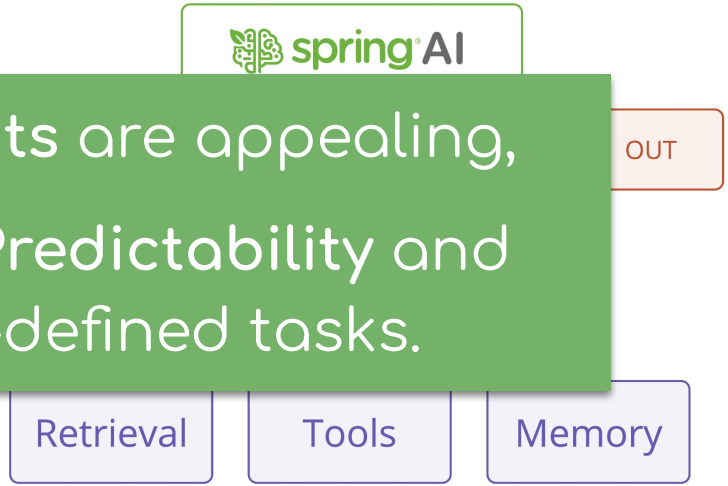
What an Agent is equipped to do
Interacts with its environment

Agentic Systems Types

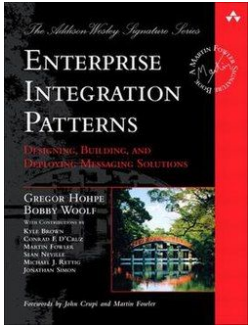
- **Workflows** - LLMs and Tools are **orchestrated through predefined** paths
- **Agent** and e toward accomplishing the tasks - e.g **autonomous**

Fully **Autonomous Agents** are appealing,

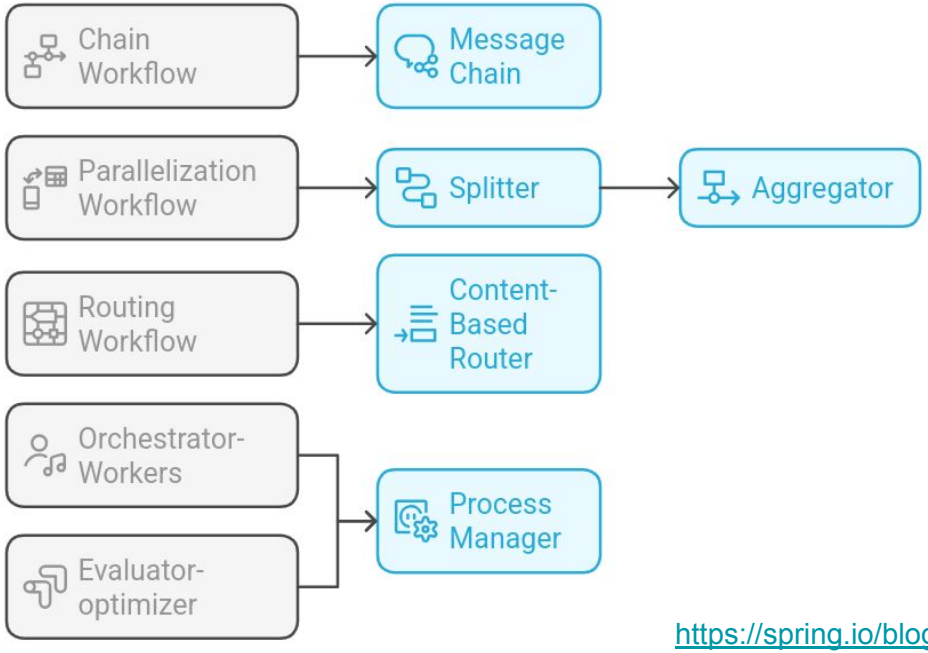
Workflows offer better **Predictability** and **Consistency** for well-defined tasks.



Agent workflow patterns

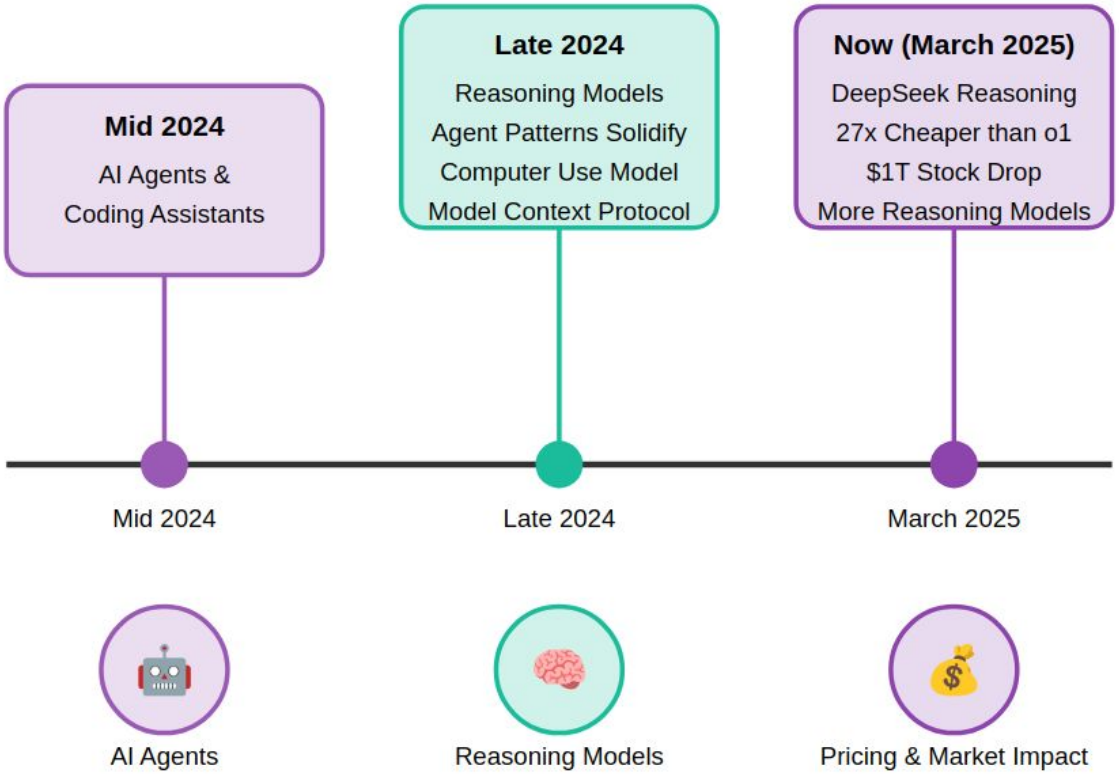


- Similar to Enterprise Integration Patterns…



https://spring.io/blog/2025/01/21/spring-ai-agentic-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