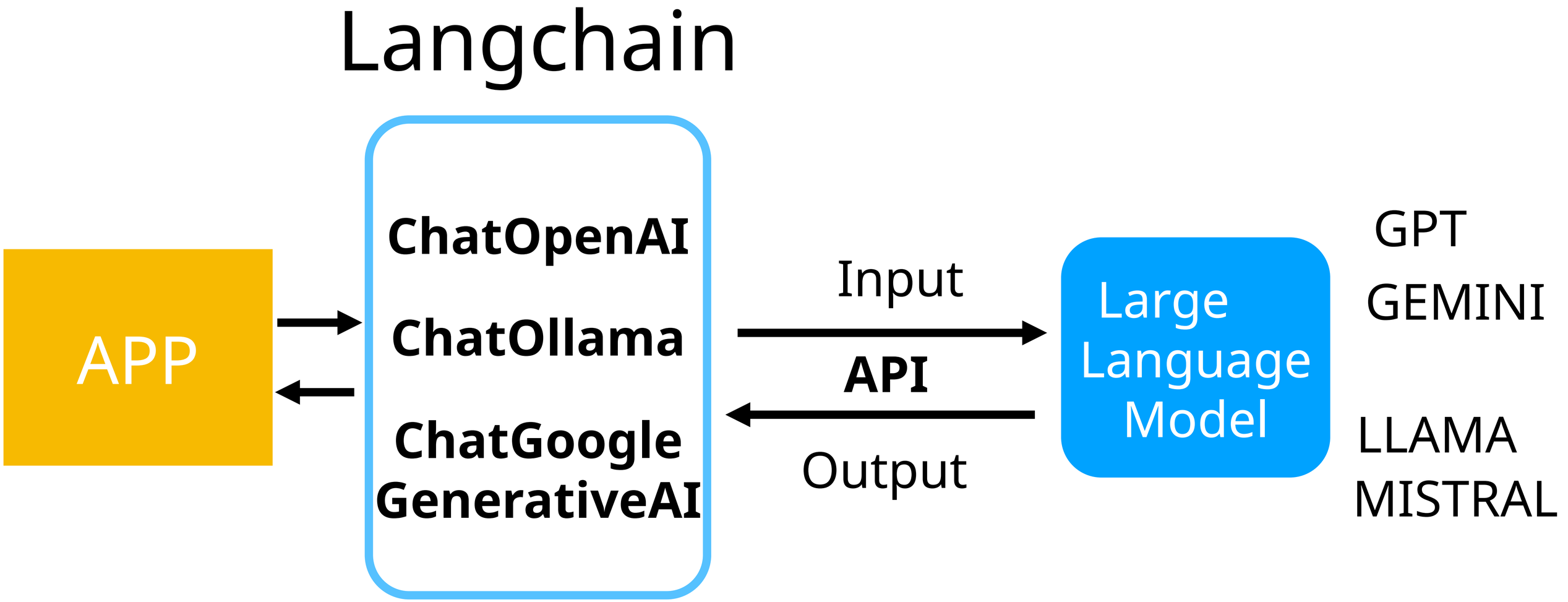
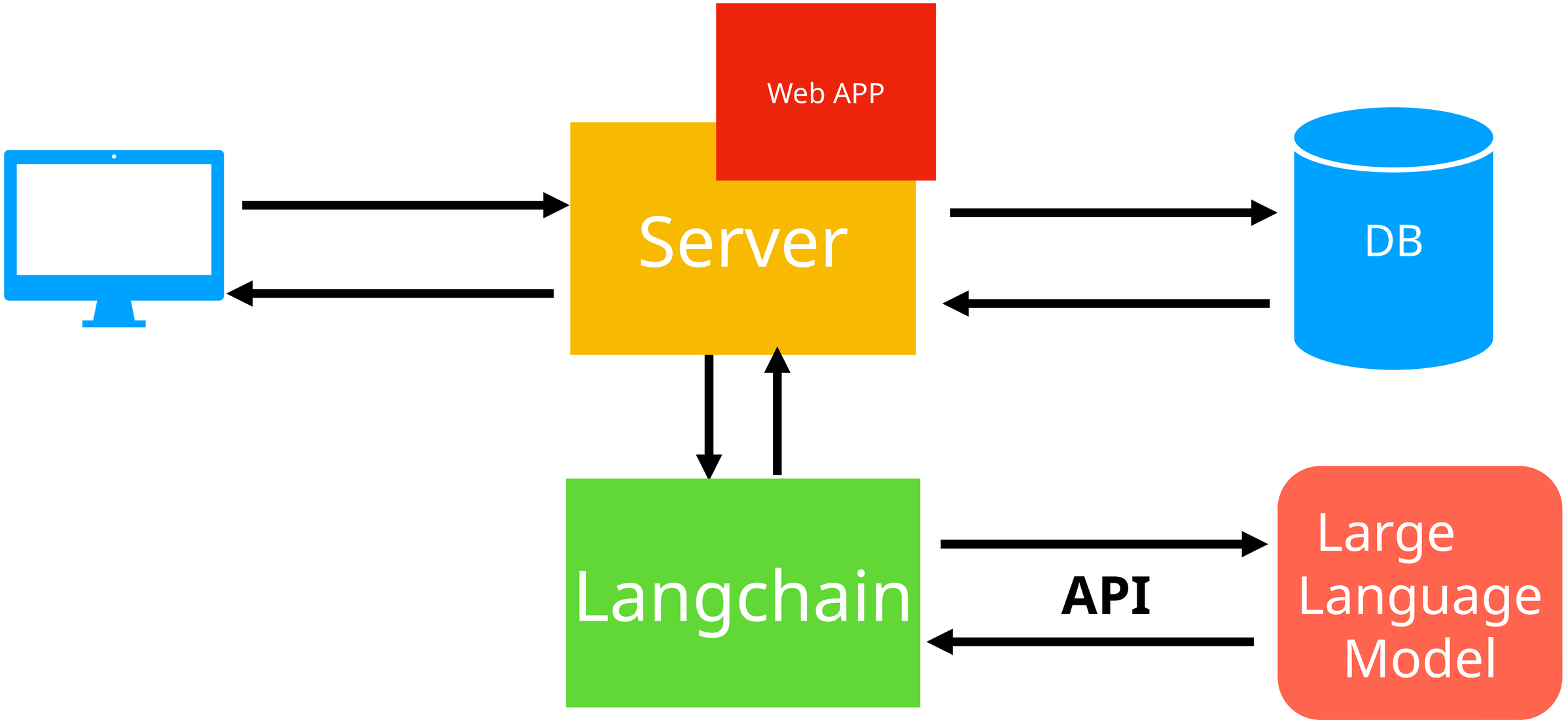
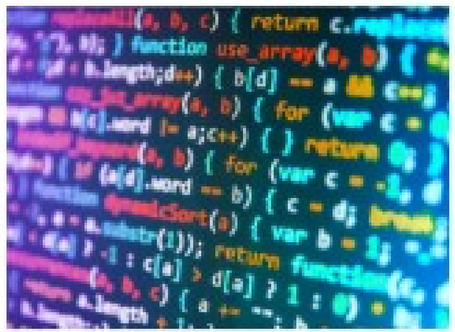

Langchain
APP
ChatOpenAI
ChatOllama
ChatGoogle GenerativeAI
Input
API
Output
Large Language Model
GPT
GEMINI
LLAMA
MISTRAL


Web APP
Server
DB
Langchain
API
Large
Language
Model

LLM



Large Language Model

Prompt

What is Machine Learning

Input

Output

Large Language Model

GPT

GEMINI

LLAMA

MISTRAL

Questions

Marketing Campaigns

Summarising Text

Assistance

Images Videos Multi Modal

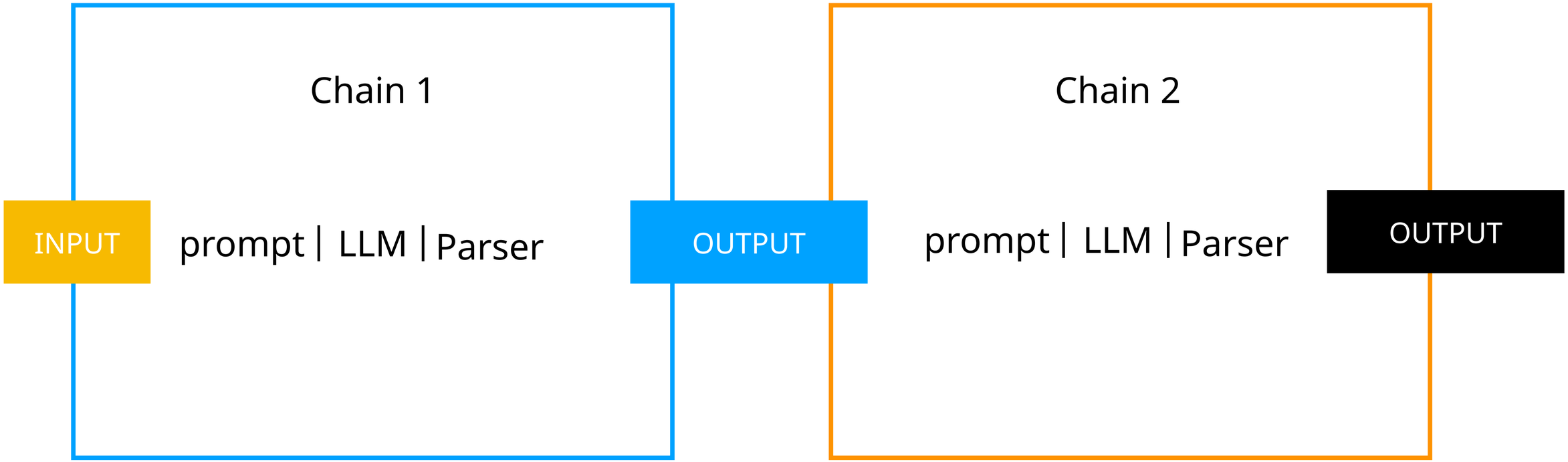
Streamlit

PromptTemplate

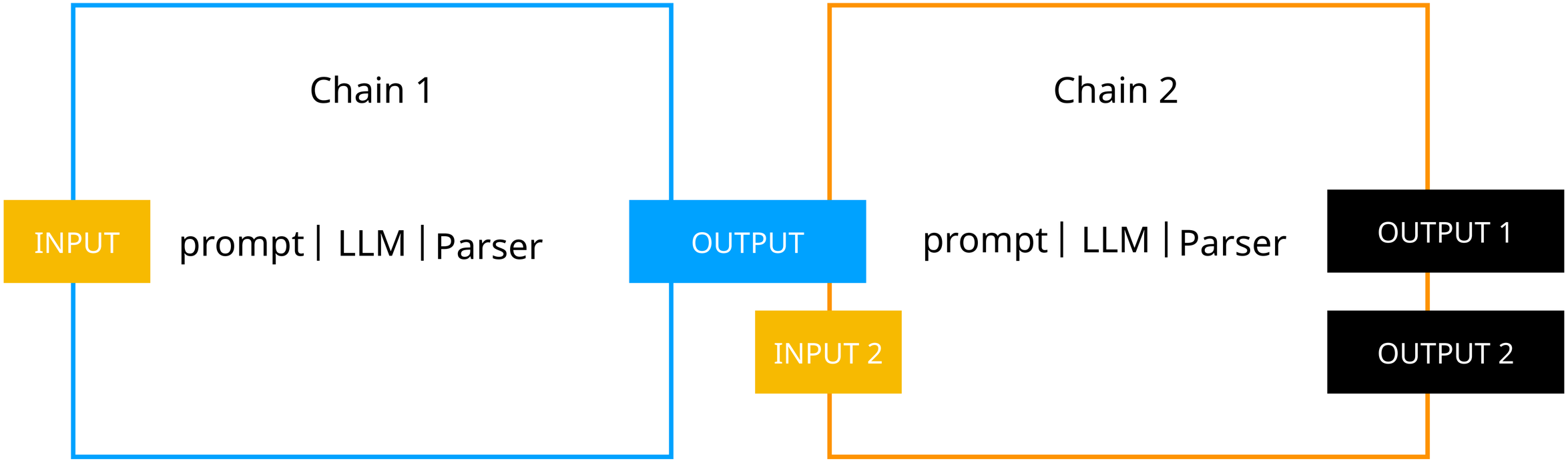
Chains

prompt | LLM

Simple Sequential Chain

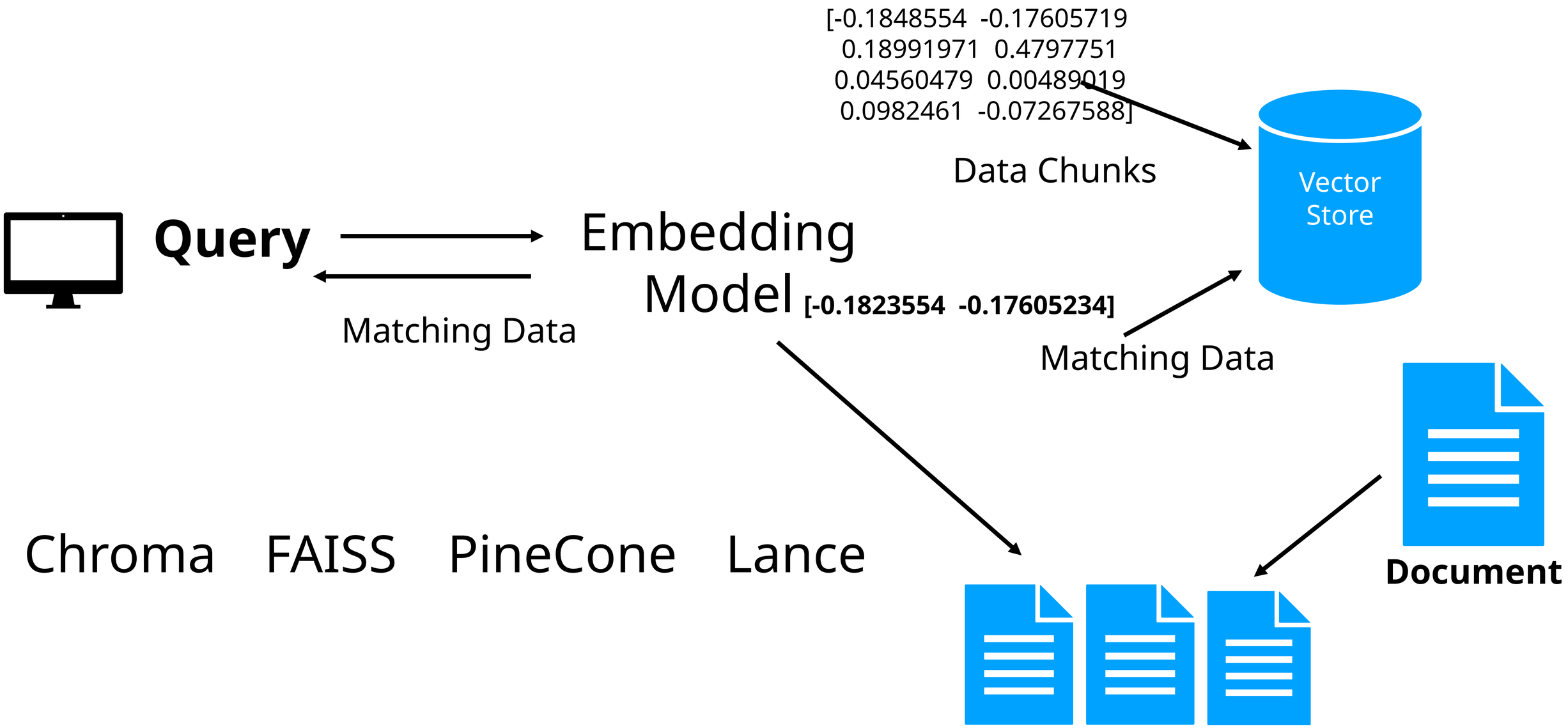


Sequential Chain

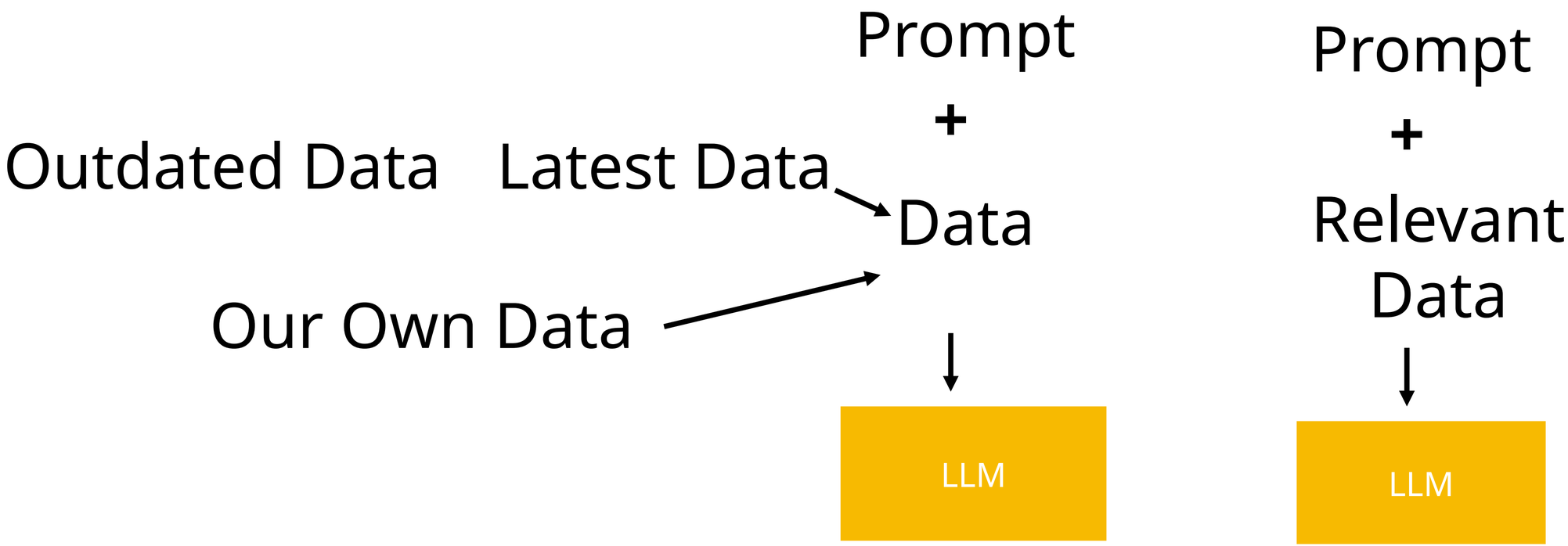


Chat History

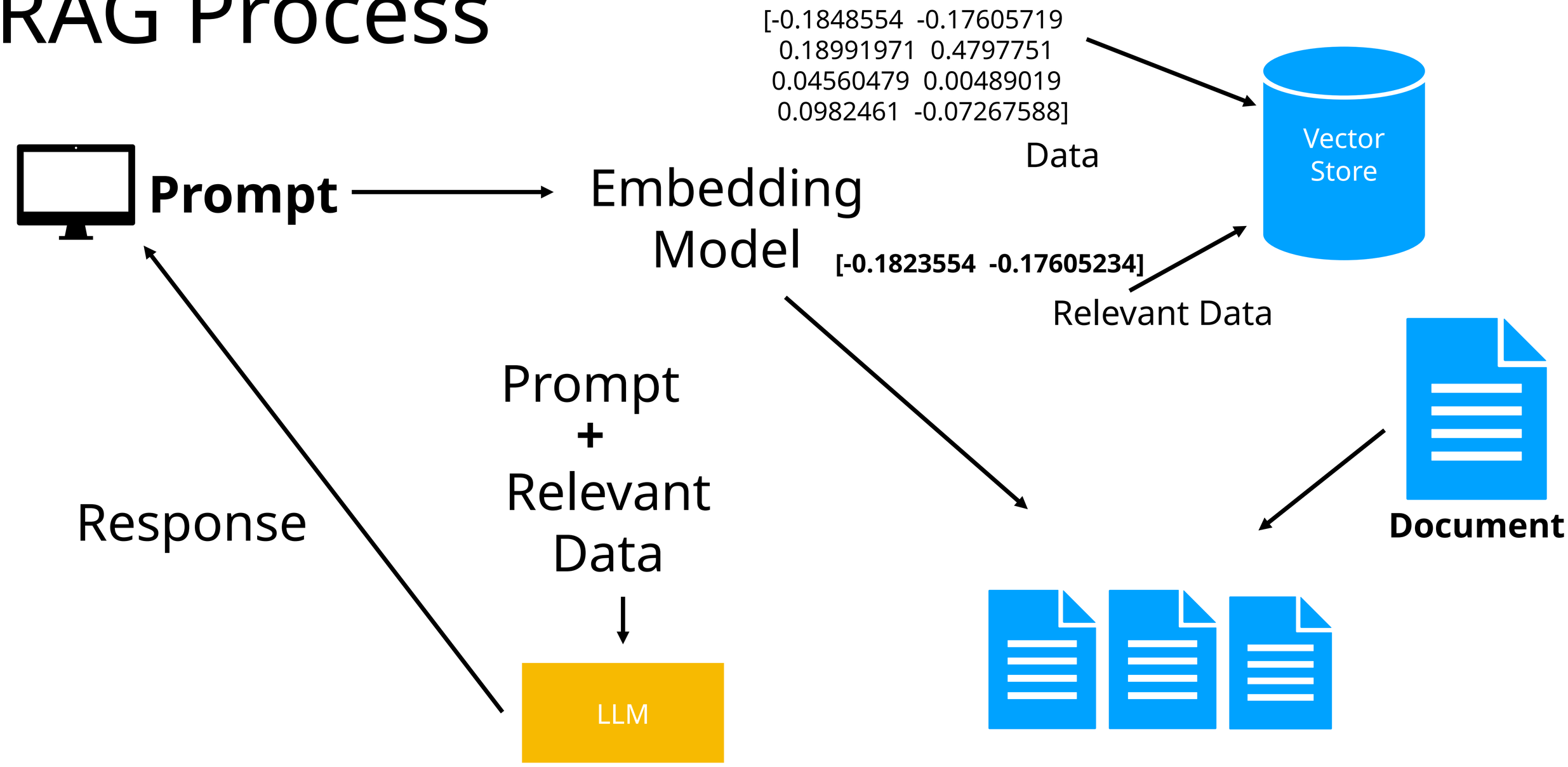
Using Vector Database



RAG Retrieval Augmented Generation



RAG Process


[-0.1848554 -0.17605719
0.18991971 0.4797751
0.04560479 0.00489019
0.0982461 -0.07267588]
Data
Vector Store
Prompt
Embedding Model
[-0.1823554 -0.17605234]
Relevant Data
Prompt
+
Relevant Data
Response
Document
LLM

Other LLMs

Google Gemini

Open Source

Llama		Gemma	Mistral
3.1	8B	2B	7B
2	7B	3B	

7B	13B	33B
8GB	16GB	32GB

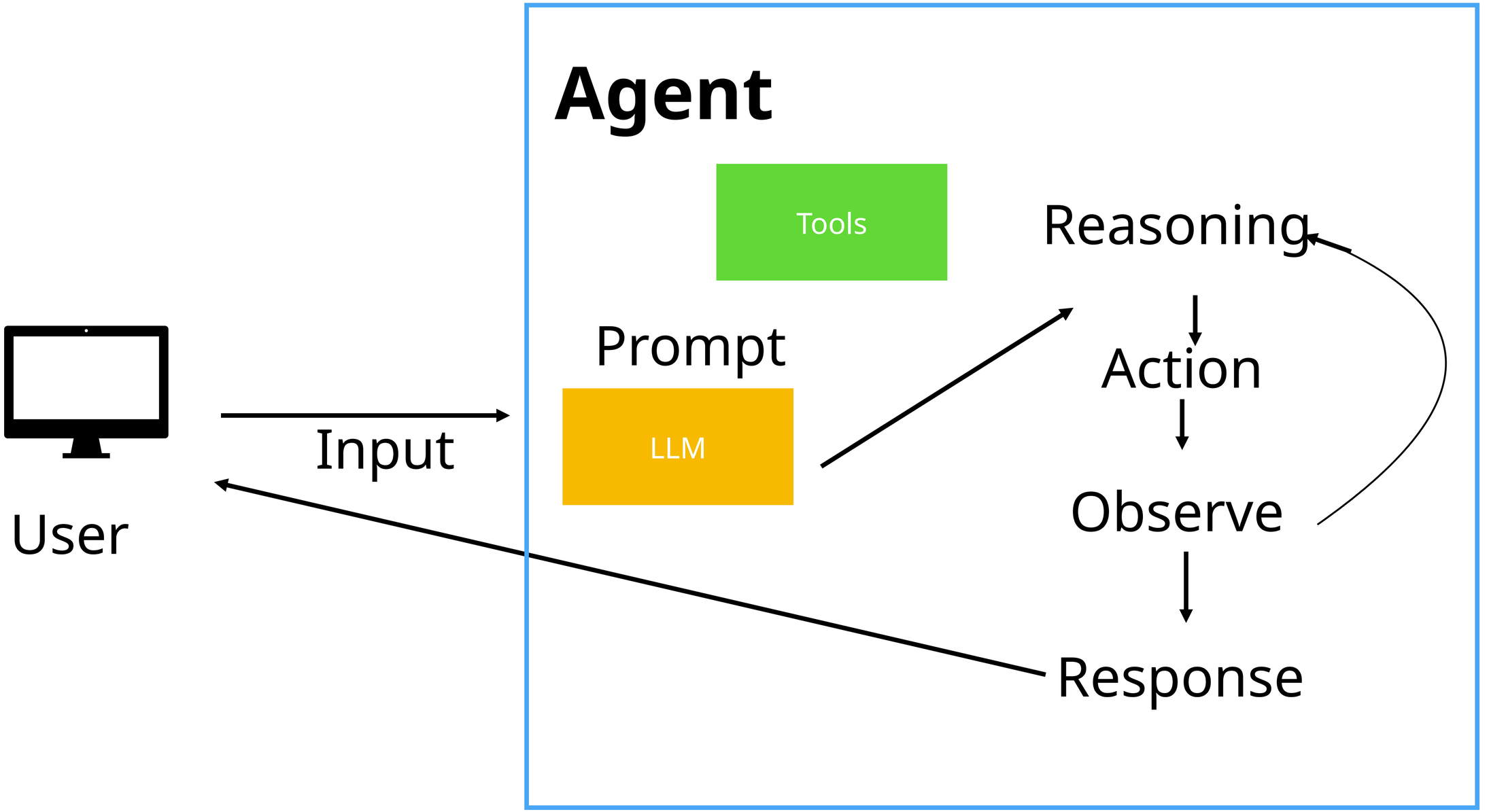
Deploy

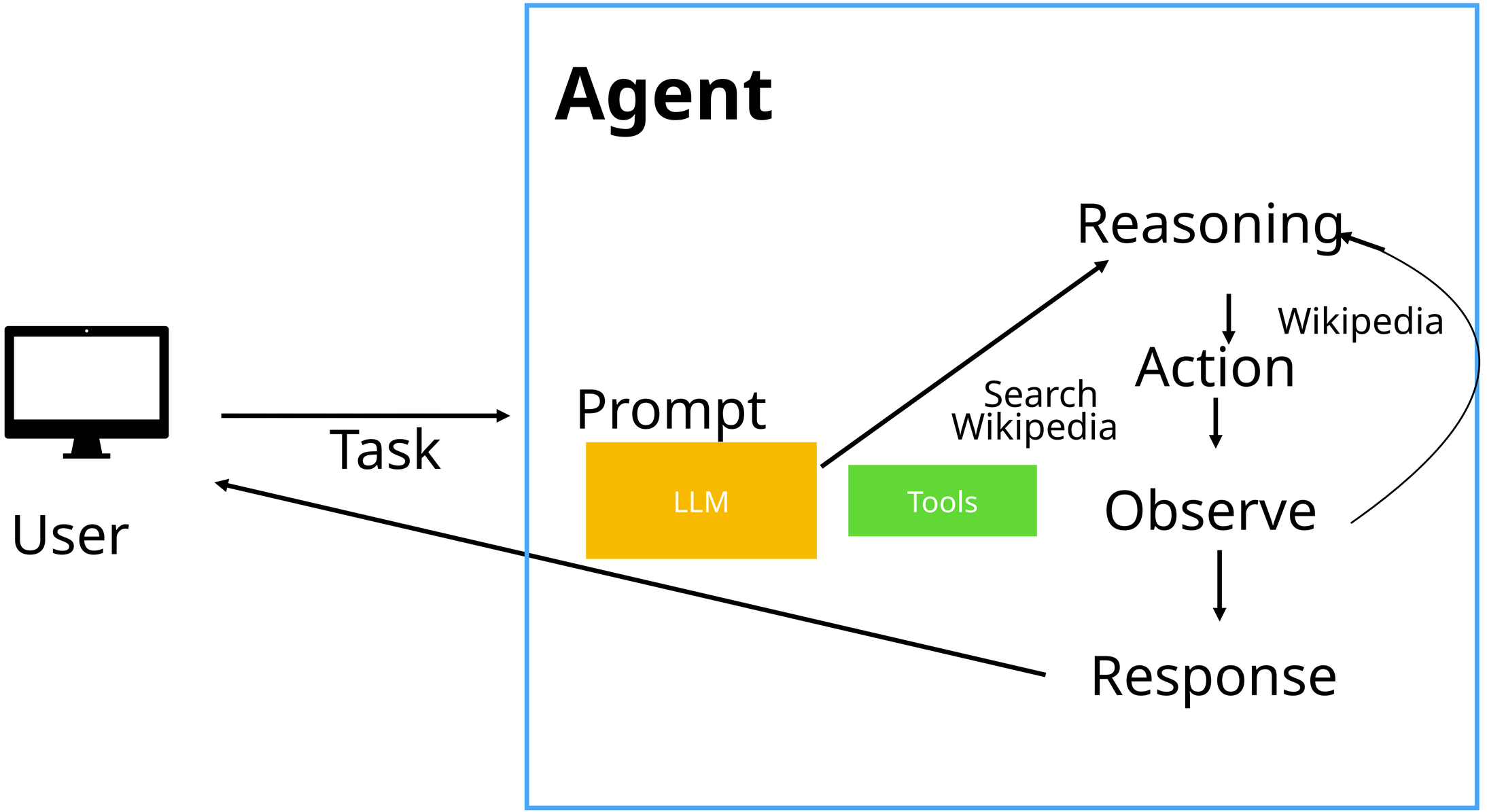
Update the code

requirements.txt

Push to github

Deploy using StreamLit Cloud

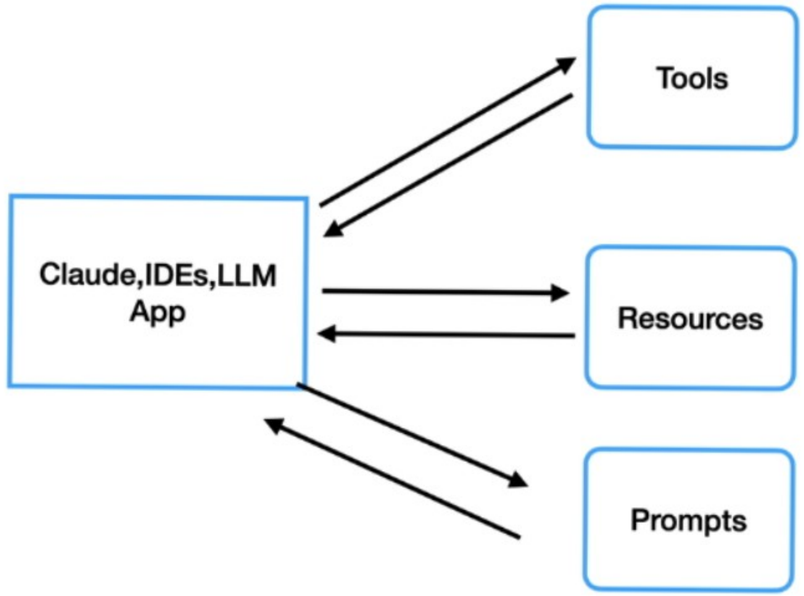

Agent
Tools
Reasoning
Prompt
Action
LLM
Observe
Input
User
Response


Agent
Reasoning
Wikipedia
Action
Search
Wikipedia
Prompt
LLM
Tools
Observe
Task
User
Response

Why MCP

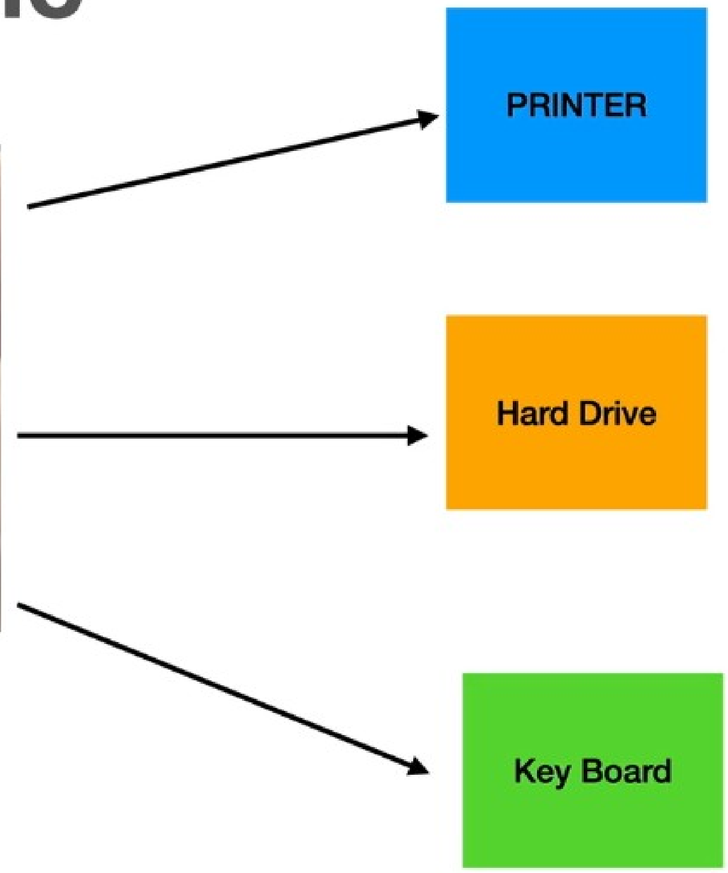
- Duplication of tools, resources and prompts per host or agents.

MCP

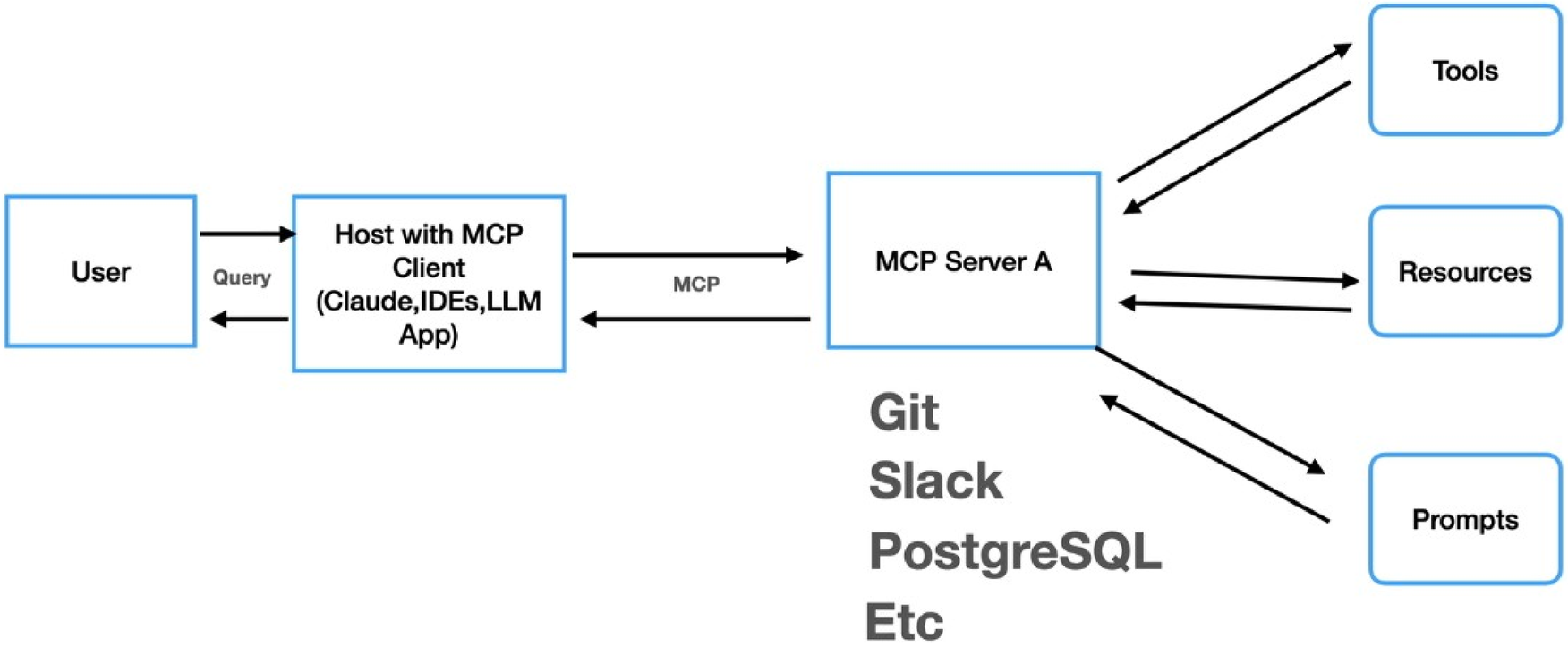


Anthropic introduced MCP

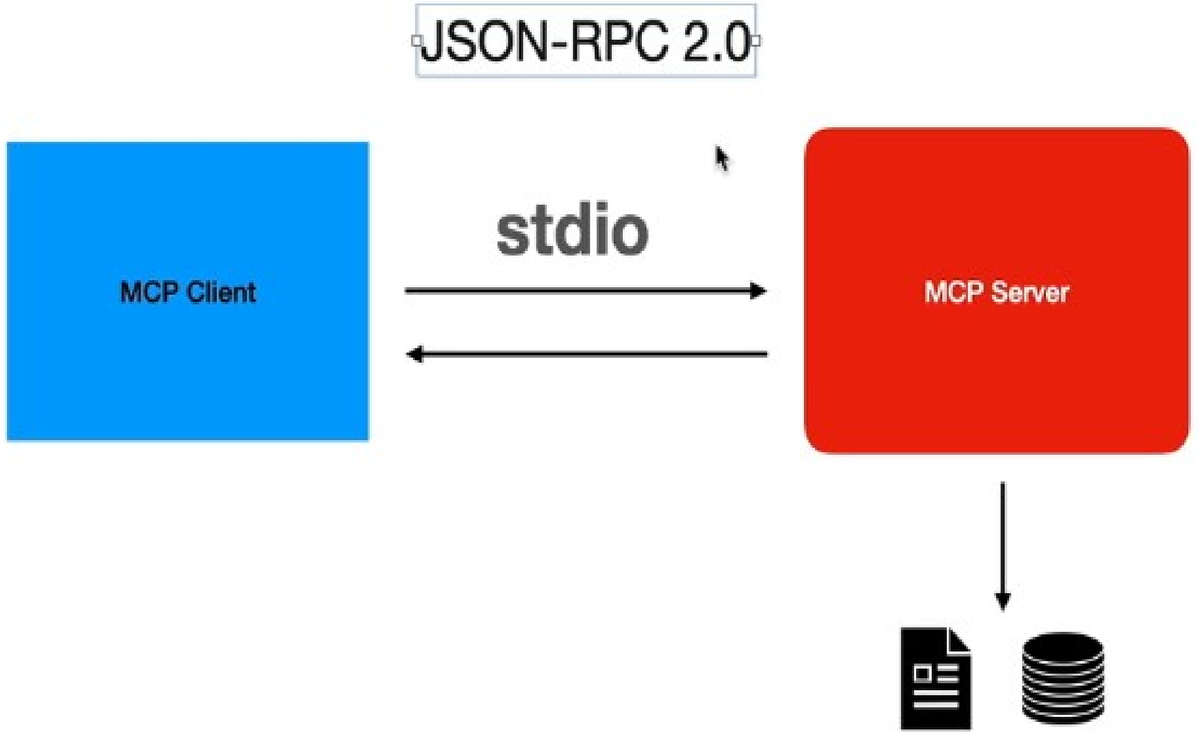
MCP Anthropic



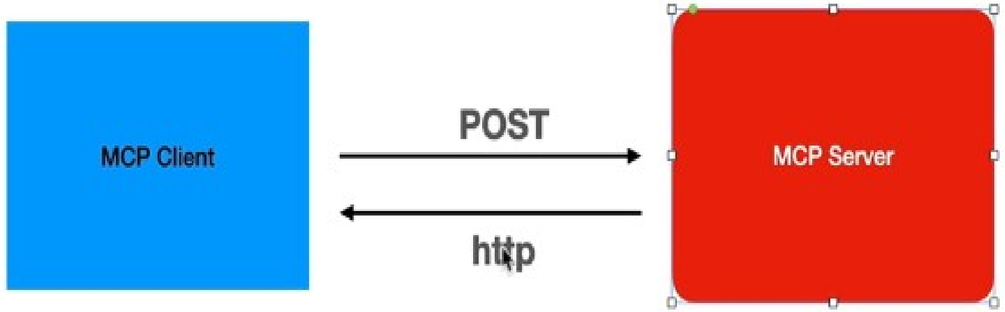
MCP


User
Query
Host with MCP Client (Claude,IDEs,LLM App)
MCP
MCP Server A
Tools
Resources
Prompts
Git
Slack
PostgreSQL
Etc

Transport



Streamable HTTP



WebSocket

SSE