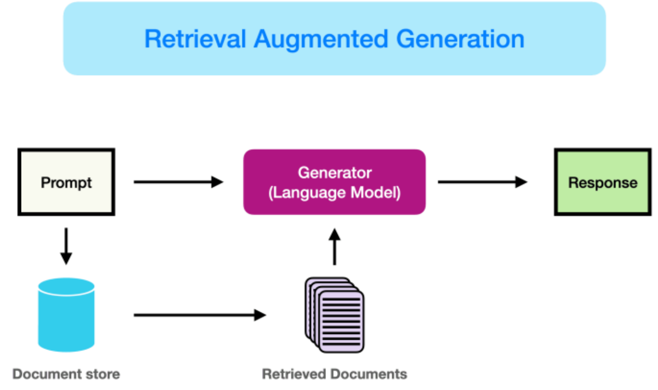
Agenda

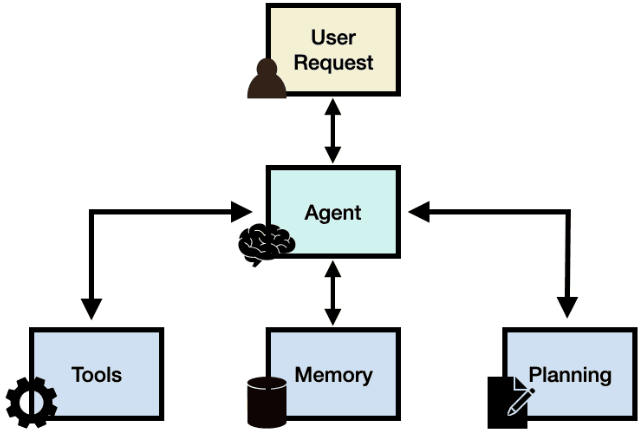
- AI Today: The Generative Landscape
- Serving a LLM On-Prem: Open WebUI
- *Coffee break*
- RAG – Knowledge Bases and Linking a SQL Database
- **How Agents Talk to Tools: Towards MCP**
- Closing Remarks

Augmenting LLM capabilities

Various options:

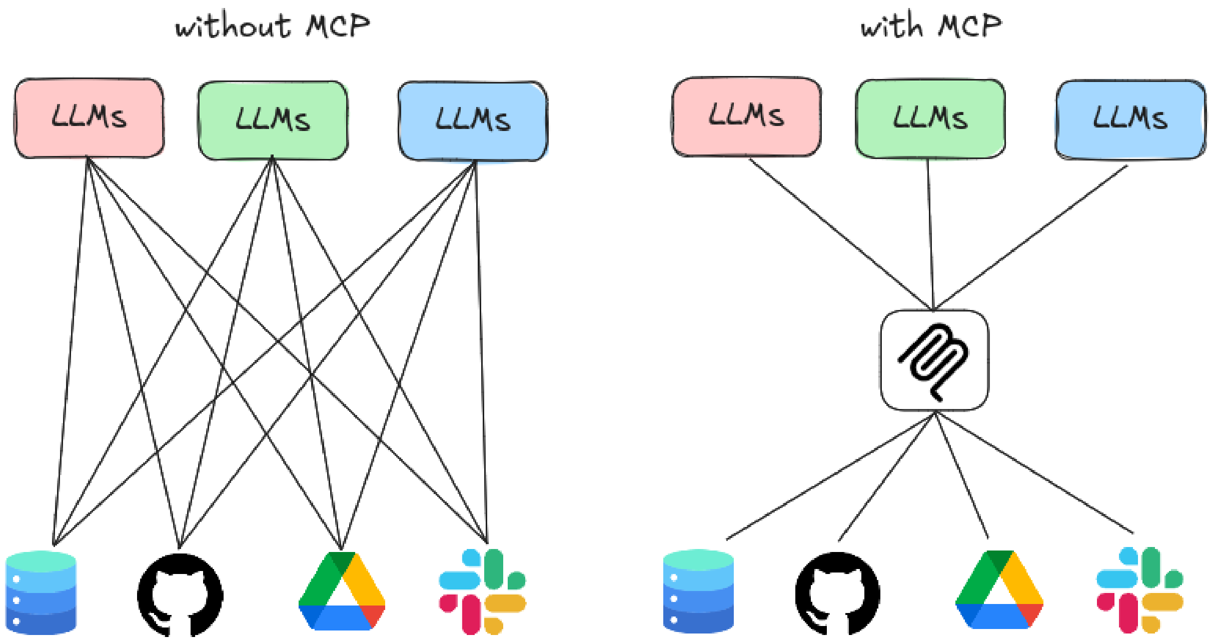
- Prompting strategies
- Retrieval-Augmented Generation (RAG)
  - Using a local knowledge base or websearch
- Agents with tools
  - Code execution, parsing of files, sending emails, …





Model Communication Protocol (MCP)

- **Advantage:** LLMs can interact with tools in a structured 2-way connection → no more “glue code” for every connection between a LLM and a tool



MCP is becoming an industry-standard

- The biggest players are adopting MCP into their products
- Also many community-built MCP servers

Microsoft
Google
OpenAI

Launching your own MCP server

- We will now set up our own MCP server

Follow the instructions on the page ***Tools and MCP -> Model Context Protocol*** to launch your own MCP server

- If you run into problems, please approach one of us!

More advanced MCP servers

- Have a look at the exercises in **Tools and MCP -> Different Tools**
  - Building MCP servers for different datasets
  - Building MCP servers with LLM- or API-powered tools

- What kind of tools for LLMs would come in handy in your daily workflow?