

Rest API

09 October 2025 09:30

allama

- ✓ → CLI
- ✓ → Python [allama]

Models → Modelfile

- Manifest
- Blobs / Layers
- Config ←

✓ CLI, Python allama

Powered by one core component

Rest API → 11434

→ Gemini,

1 → LLM → Server → API Address

- ✓
- ① Chating
- ② Development

Chating

2 → Chatbot → login [Google] → credentials

Model → Gem 2.5 Pro

Google Gemini

Wrapper → Rest API

③ → User → Prompt → Rest API [Brains] → Response

Development

→ API → Key. ✓

Rest [Brain]

Development

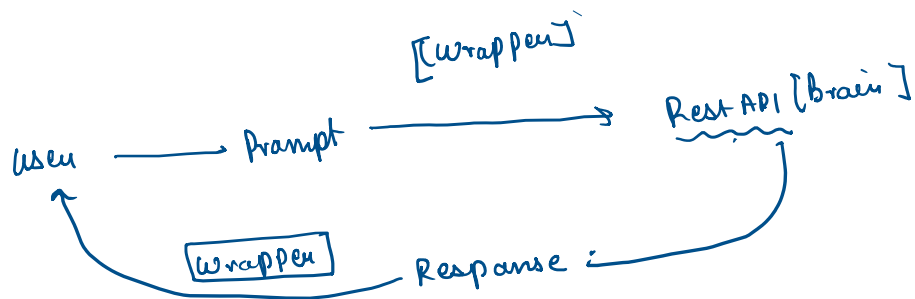
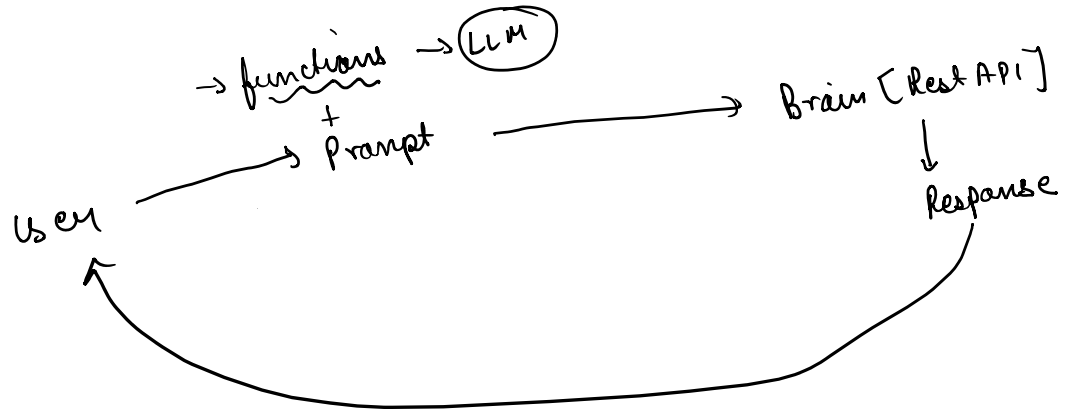
→ API → Key ✓

→ library module
↳ wrapper

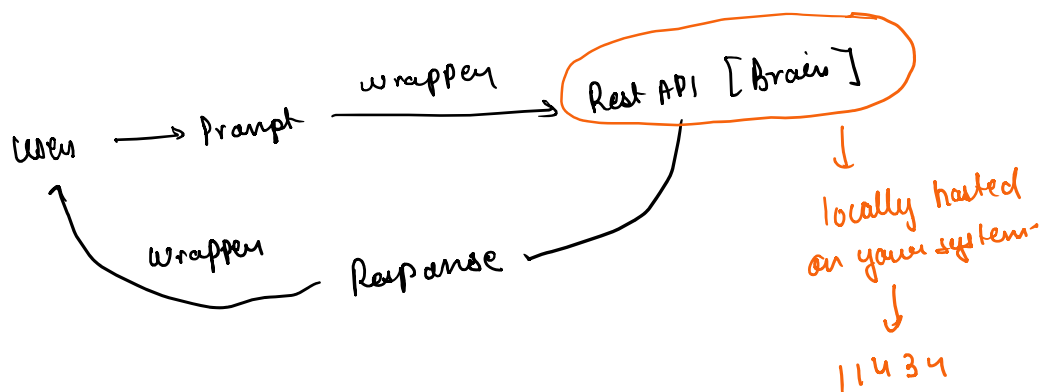
→ Rest API [Brain]

Rest [Brain]

ollama-gen (cli)



ollama follows the same Rest API Architecture



CLI → ollama run llama3.2:1b "why is sky blue?"
or
prompt = "why is sky blue?"

CLI → ✓ ollama run llama3.2:1b

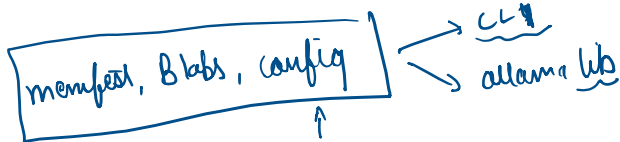
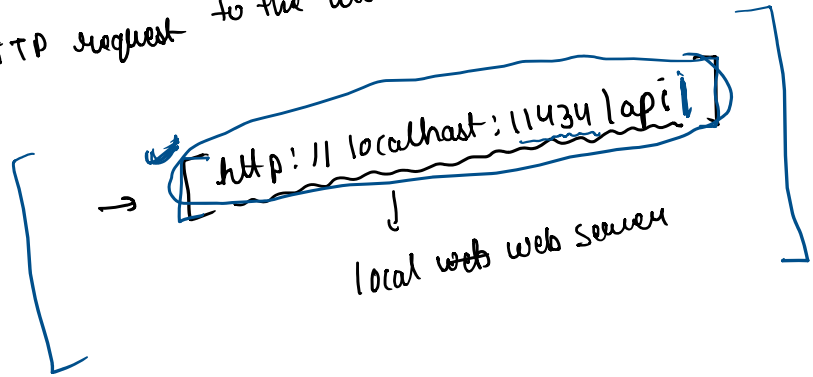
or

ollama

✓ ollama.generate(model = _____, prompt = "with a sky blue")

✗ directly talking to llama3.2:1b

↓
HTTP request to the web address



→ ollama pull < >
↳