

Generative AI

LLM → Limited power

GPT 3.5 Turbo → 2022 October

Many Providers:

- ① OpenAI ←
- ② Llama ← Llama ^{CP} LLMs
- ③ Mistral ← SDK
- ④ Falcon ← SDK
- So on - - - -

problem:

- * Limited context window
- * Outdated knowledge base
- * Data privacy
- * Cost
- * No connection with 3rd party tool
- So on - - - -

Gen AI → APPS

Data in:

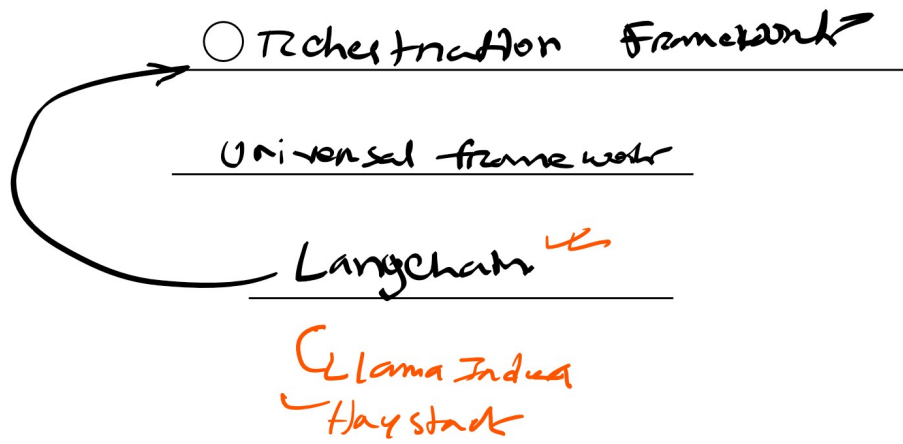
- txt
- pdf
- CSV
- SQL
- xls
- html
- png
- So on - - -

Action

- ↳ RAG
- Agents
- APIs
- format
- manipulate
- Translation
- memory
- 3rd party tools
- So on - - -

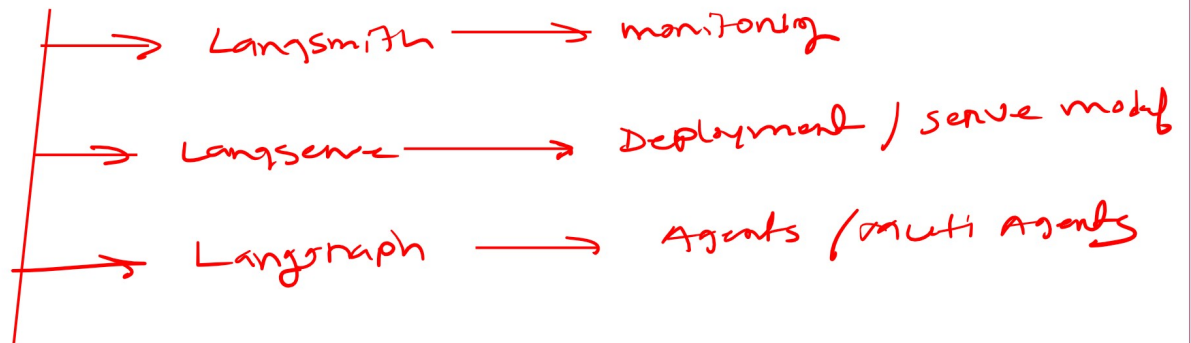
Data out:

anything



Why we chose?

Langchain: Orchestration Framework




Langchain Evolution:

- * From LLM model to chat model
- * From Legacy to LCEL - Langchain Expression Language
- * From LC to LC Ecosystem
- * From terrible Docs to - - - - -

Generative AI space is the fastest evolving space in software. No matter how hard I work the content will never be 100% up to date.

Note: But you will be ready: →

I do not give you fish
I teach you how to fish



① Explore —

② Research —

③ practice —

Learning path

✓ ① Introduction.

✓ ② How to talk with LLM (OpenAI, OpenSource)

③ how to work with data — Basic RAG

④ Chain | Pipelines

⑤ memory

⑥ Longchain Ecosystem

→ Langserve
→ Langsmith
→ Langgraph

⑦ Level -1 APPS → Exp

⑧ Level -2 APPS → POC

⑨ Level -3 APPS → professional