

Day 22: LangChain for LLMs

Concept

Intro

- Open-source frame work
- We already understood it's application in the last session about RAG
- It helps us create applications using LLMs
- It helps 'chain' LLMs with external computation and data

Pipeline

- User asks a question to the language model
- Similarity search is performed in the vector DB
- Relevant info is fetched
- Relevant info + original query passed as the prompt to the language model

Modules



Models

- The LLMs
- Consist of neural networks with numerous parameters
- Trained on extensive amounts of data
- Eg: BERT, GPT, LaMDA, PaLM, LLaMa

Prompt

- prompts that are dynamically generated based on user input
- Once you have a good prompt, use it as a template for other purposes

Memory

- way of keeping track of the state of a conversation
- We can provide context from previous messages

Chains

- To combine multiple components together to solve a specific task

Agents

- Allow LLMs to interact with their environment

Hands-on