LLM – Large Language Models

Lang Chain – It is an open source library that provides developers with tools to build applications powered by large language models

Chain together different components to create more advanced use cases along LLM

Models – You choose the model

Prompts – Give in the question [Formatting the Prompt]

Memory – Make sure the bot doesn’t forget the previous user interaction. [Conversation chain]

Indexes – Language models are more powerful when combined with our own text data.

Chains – Sequence of calls, Lots of integration with other tools

Agents – Makes decisions about which actions to take, taking that action, seeing an observation and repeating that until done

[Langchain library has document loaders, Text splitters and vector stores]

Document loaders are little helper tools that make it easy to download documents

The data is too large to send inside the model which is not possible So, we use the text splitters (Max tokens are 4096, 4097 for the top performing models)

Gpt 4 has the maximum of 32768 tokens

Faiss library developed by Meta used for efficient Similarity search

docs= db.similarity\_search(query, k=k)

the k parameter defaults to 4.