

LLM RoadMap

Langchain

- What is Langchain
- Why langchain?

ChatGpt and Langchain Intergration

- setting up Python environment
- Creating APIKeys
- Using Langchain
- Intoduction to Chains and type of chains
- Using Chains
- Sequential Chains

Memory

- Type of users: HumanMessage, userMessage, AiMessage
- What is Memory and Why we need Memory
- Using chat/conversational Chains
- Chat chains vs Completion Chains
- Understanding Memory
- Types of Memory
 - ConversationalBufferMemory
 - CombinedMemory
 - ConversationTokenBufferMemory
 - ConversationBufferWindowMemory
 - Saving and Extending Conversations
 - Json
 - Databases
 - ConversationSummaryMemory

Embeddings

- Loading Files with document loaders
- Introduction to embeddings
- Chunking Text
- Generating Embeddings

Document Retrievers

- introduction to ChromaDB
- Retrieval chain
- Retriever
 - stuff
 - refine
 - MapReduce
 - Map Rerank
- Visualizing Embeddings

Tools and Agents

- Understand Tools and why tools
- Understanding ChatGPT functions
- Defining a tool
- Define an Agent and Agent executor
- Tools error handling
- Tools with arguments
- Memory vas Agent ScratchPad
- Callbacks and Callbacks implementation

Pinecone

- Load and Spliting PDF
- Introduction to Pinecone
- Initialize pinecone Client
- Adding documents to the Vector stores
- Background Services
 - Celery
 - Nodecron

Custom Message Histories

- Memory vs SQL database
- Building a retriever
 - ConversationalRetrieverChain
- Storing Messages in SQL database

Streaming

- Experiement with a streaming language Model
- Chains don't stream Issue
- using callbacks

Self-Improving Text generation

- Collecting user feedback
- Altenative answers
- LLM Scores

LangSmith

- How to monitor aps

LangServe

- Deployment