**Prompt Engineering:**

Prompt Engineering & basic of LLM

* What is difference between Predictive/ Discriminative AI and generative AI?

Free Preview

View Lesson

* What is LLM & how LLMs are trained?

Free Preview

View Lesson

* What is token in language model?

View Lesson

* how to estimate the cost of running a SaaS-based & Open source LLM model?

View Lesson

* Explain Temperature parameter and how to set it?

View Lesson

* What are different decoding strategies for picking output token?

View Lesson

* What are different ways you can define stopping criteria in large language model?

View Lesson

* How to use stop sequence in LLMs?

View Lesson

* Explain basic structure of prompt engineering?

View Lesson

* Explain In-Context learning?

View Lesson

* Explain type of prompt engineering?

View Lesson

* What are some of the aspect to keep in mind while using few-shots prompting?

View Lesson

* What are certain strategies to write good prompt?

View Lesson

* What is hallucination & how can it be controlled using prompt engineering?

View Lesson

* How do I improve reasoning ability of my LLM through prompt engineering ?

View Lesson

* How to improve LLM reasoning if your COT prompt fails?

**RAG:**

Retrieval augmented generation (RAG) Systems

6 lesson(s)

* how to increase accuracy, and reliability & make answers verifiable in LLM

View Lesson

* How does Retrieval augmented generation (RAG) work?

View Lesson

* What are some of the benefits of using RAG system?

View Lesson

* What are architecture patterns you see when you want to customize your LLM with proprietary data?

View Lesson

* When should I use Fine-tuning instead of RAG?

**Data chunking:**

Data Chunking

5 lesson(s)

* What is chunking and why do we chunk our data?

View Lesson

* What are factors influences chunk size?

View Lesson

* What are different types of chunking methods available?

View Lesson

* How to find ideal chunk size?

**Embedding:**

Embedding Model

7 lesson(s)

* What are vector embeddings? And what is embedding model?

View Lesson

* How embedding model is used in the context of LLM application?

View Lesson

* What is difference in embedding of short and long content?

View Lesson

* How to benchmark embedding models on your own data?

View Lesson

* Scenario based question

View Lesson

* Walk me through steps of improving sentence transformer model used for embedding?

**Vector DB:**

* What is vector DB?

View Lesson

* How vector DB is different from traditional database?

View Lesson

* How does a vector database work?

View Lesson

* Explain difference between vector index, vector DB & vector plugins?

View Lesson

* What is different vector search strategies?

View Lesson

* Scenario based question

View Lesson

* How does clustering reduce search space? When does it fail and how can we mitigate these failures?

View Lesson

* Explain Random projection index?

View Lesson

* Explain Locality-sensitive hashing (LHS) indexing method?

View Lesson

* Explain product quantization (PQ) indexing method?

View Lesson

* Compare different Vector index and given a scenario, which vector index you would use for a project?

View Lesson

* How would you decide ideal search similarity metrics for the use case?

View Lesson

* Explain different types and challenges associated with filtering in vector DB?

View Lesson

* How do you determine the best vector database for your needs?

**Search:**

* Why it’s important to have very good search

View Lesson

* What are the architecture patterns for information retrieval & semantic search, and their use cases?

View Lesson

* How can you achieve efficient and accurate search results in large scale datasets?

View Lesson

* Scenario based question

View Lesson

* Explain the keyword-based retrieval method

View Lesson

* How to fine-tune re-ranking models?

View Lesson

* Explain most common metric used in information retrieval and when it fails?

View Lesson

* Scenario based question

View Lesson

* I have a recommendation system, which metric should I use to evaluate the system?

View Lesson

* Compare different information retrieval metrics and which one to use when?

**Large Language Models:**

* Can you provide a detailed explanation of the concept of self-attention?

View Lesson

* Explain the disadvantages of the self-attention mechanism and how can you overcome it.

View Lesson

* What is positional encoding?

View Lesson

* Explain Transformer architecture in detail?

View Lesson

* What are some of the advantages of using a transformer instead of LSTM?

View Lesson

* What is the difference between local attention and global attention?

View Lesson

* What makes transformers heavy on computation and memory, and how can we address this?

View Lesson

* How can you increase the context length of an LLM?

View Lesson

* If I have a vocabulary of 100K words/tokens, how can I optimize transformer architecture?

**Supervise Fine-tuning (SFT) LLM:**

* What is finetuning and why it's needed in LLM?

View Lesson

* which scenario do we need to finetune LLM?

View Lesson

* How to make the decision of fine-tuning?

View Lesson

* How do you create a fine-tuning dataset for Q&A?

View Lesson

* How do you improve the model to answer only if there is sufficient context for doing so?

View Lesson

* How to set hyperparameter for fine-tuning

View Lesson

* How to estimate infra requirements for fine-tuning LLM?

View Lesson

* How do you fine-tune LLM on consumer hardware?

View Lesson

* What are the different categories of the PEFT method?

View Lesson

* Explain different re-parameterized methods for fine-tuning LLM

**Hallucination:**

* What are the different forms of hallucinations?

View Lesson

* How do you control hallucinations at different levels?

**Evaluation:**

* How do you evaluate the best LLM model for your use case?

View Lesson

* How to evaluate RAG-based system

View Lesson

* What are the different metrics that can be used to evaluate LLM

View Lesson

* Explain the Chain of verification

View Lesson

* Test yourself

**Agents:**

* Explain the basic concepts of an agent and the types of strategies available to implement agents

View Lesson

* Why do we need agents and what are some common strategies to implement agents?

View Lesson

* Explain ReAct prompting with a code example and its advantages

View Lesson

* Explain Plan and Execute prompting strategy

View Lesson

* Explain OpenAI functions strategy with code examples

View Lesson

* Explain the difference between OpenAI functions vs LangChain Agents