GEN-AI MASTERS PROGRAM

By Sumit Mittal

Getting started with Gen AI

- Introduction to GenAI and LLMs
- Pre training and Fine Tuning
- Real-time Support Chatbot use-case
- RAG, RLHF, Longchain, Few Shot Learning
- Llama model

Internals of Transformers

- Transformers Architecture
- How Transformers Work
- Token Embeddings
- Tokenization, Embeddings, Context

Gen-AI Practicals

- Introduction to huggingface - Downloading a Model from HuggingFace | Execute NLP Tasks (Sentiment Analysis)
- Text Summarization | Translation | Question - Answer Tasks
- Token Classification | Fill Mask | Text Generation
- Feature Extraction | Zero Shot Classification
- Selecting the Right Model for the Task | Pre-Processing
- Model Inference | Working of Softmax
- More about Tokenizer - Auto Tokenizer | Auto Model

Transformers Internals | Attention Blocks

- Model selection
- Preprocessing - tokenizer
- Postprocessing - SoftMax
- Attention Block and Multilayer Perceptron : Decoding Attention Pattern
- Attention Pattern : Vectors and Matrices
- Embedding matrix & Unembedding matrix
- Query matrix & Query vector | Key vector & Key matrix
- Attention Mechanism : Embedding the Contextual Knowledge
- Multi Layer Perceptron : Feed Forward Layer
- How MLP Works

Training a model

- Introduction to the Bigram Model | Understanding tensors
- Bigram Model
- Tensors
- Pretraining with Matrices
- Word Generation
- Precomputed Probabilities
- Model Quality
- One hot Encoding and Forward Pass
- Backward Pass and Adjusting Weights

Model Pre-Training

- Implementation of Multilayer Perceptron
- Probability Calculation
- Tuning Parameters
- Optimized Industrial Approach - Mini Batches

Getting Started with RAG

- Introduction to RAG
- RAG Practicals (Using Google Colab)
- RAG Practicals Using Databricks CE & Local
- Accessing Gated Models
- RAG Internals Explained
- RAG Pipeline : Text to Embeddings
- RAG Pipeline : Store Embeddings & Retrieve Answers with LLM

RAG End-to-End Production Pipeline

- End-to-End Production Grade Pipeline
- Generating Embeddings
- Building Retriever and RAG LangChain Creation
- Registering Lang-chain and Creating Serving Endpoint

LangChain Essentials

- Understanding LangChain Framework
- Inferring Large Models on Cloud
- Working with OpenAI
- Message Structure : System | Human | AI Message
- Usecase : Few Shot Learning
- Prompt Template
- Task & Chain : Runnable Lambda & Runnable Sequence
- Runnable Parallel

RAG Application using LangChain

- Advanced RAG Application Use-case with Langchain
- Document Loaders
- Chunking Strategies & Embedding Model Selection
- Retriever Configs | Search Types
- Conversational RAG Solution
- RAG Challenges

Query Optimization

- Query Transformation
- Query Routing
- Indexing Strategies
- Tracing and debugging with LangSmith
- More on Query Transformation Techniques - Query Rewrite
- Query Optimization Techniques - Multi Query
- Rag Fusion | Reciprocal Ranking
- Query Decomposition | Sub-Query | Chain of Thoughts(COT)
- Query Decomposition | Multiple Independent Sub-Queries
- Step Back Questions | Few Shot Prompting

AGENTIC AI

- Agentic Behavior in AI
- Introduction to LangGraph
- LangGraph Use Case: Natural Language to PySpark DataFrame
- LangGraph Use Case: PySpark DataFrame to Spark SQL
- Binding Tools to LLMs Using LangGraph

AGENTIC AI TOOLS

- Registering Tools with LLMs in LangGraph
- Tool Binding with ReAct Architecture
- State Persistence in LangGraph Using Checkpoints
- Structured Outputs with Pydantic Models | Reducers to Retain Full Conversation History
- Handling Long Conversations with Message Filtering
- Dynamic Summarization in LangGraph Agents
- Persisting LangGraph State with SQLite Checkpointers

CRAG | AGENTIC AI PROJECTS

- Corrective RAG (CRAG)
- Building a LangGraph based CRAG Application
- Agentic AI Project - SQL Querying Agent with Tools
- Designing a Multi-Role LLM System with Agents

LLM Fine-Tuning

- Introduction to LLM Fine-tuning
- RAG Vs Fine-Tuning
- Types of Fine-Tuning
- Full Fine-Tuning & Parameter Efficient Fine-Tuning (PEFT)
- LoRA | QLoRA (Reducing the Memory & Compute Requirement)
- Practicals Fine-Tuning Demo
- Hyperparameters

Thank You

Website: https://trendytech.in/

TRENDYTECH
UPLIFT YOUR CAREER!