

PART: 2

APPLIED GEN AI

Module 1: Introduction to LLMs and Fundamentals

Introduction to LLMs and Fundamentals

To build a foundational understanding of what LLMs are, their history, and their basic architecture.

Week 1-2: Begin with the basics.

- Understand what Large Language Models (LLMs) are.
- Familiarize yourself with key LLM terminologies.
- Explore the evolution from language models to LLMs.
- Look into the history of Natural Language Processing (NLP) and LLMs.

Module 1: Introduction to LLMs and Fundamentals

Week 3-4: Dive into LLM Architectures

- Study the Transformer architecture.
- Examine Transformer model design choices.
- Learn about the Generative Pre-trained Transformer (GPT) architecture.
- Get an introduction to Large Multimodal Models.
- Understand the differences between proprietary, open, and open-source language models.

Module 2: LLMs in Practice and Prompting

LLMs in Practice and Prompting

Understand the practical aspects of working with LLMs, including their limitations and how to control their outputs.

Week 5-6: Practical aspects of using LLMs.

- Learn about potential issues like hallucinations and biases in LLMs.
- Study methods for evaluating LLM performance.
- Explore how to control LLM outputs.
- Get introduced to pretraining and fine-tuning of LLMs.

Module 2: LLMs in Practice and Prompting

Week 7-8: Learn about prompting techniques.

- Understand the basics of prompting and prompt engineering.
- Explore different prompting techniques.
- Identify bad prompt practices.
- Learn tips for effective prompt engineering.

Module 3: Introduction to LangChain and LlamaIndex

Introduction to LangChain and LlamaIndex

Introduce two key frameworks for building LLM-powered applications, LangChain and LlamaIndex.

Week 9-10: Introduction to LangChain

- Learn the basics of LangChain.
- Explore LangChain agents and tools.
- Understand how to build LLM-powered applications with LangChain.
- Start building a simple application (e.g., a news article summarizer).

Week 11-12: Introduction to LlamaIndex

- Learn about LlamaIndex.
- Understand the differences between LangChain, LlamaIndex, and OpenAI Assistants.

Module 4: Prompting and Retrieval-Augmented Generation (RAG)

Prompting and Retrieval-Augmented Generation (RAG)

Delve into advanced prompting techniques and learn about Retrieval-Augmented Generation (RAG).

Week 13-14: Advanced Prompting with LangChain.

- Learn about LangChain Prompt Templates.
- Study few-shot prompts and example selectors.
- Explore output parsers for managing LLM outputs.
- Improve the news article summarizer application.
- Learn how to create knowledge graphs from textual data.

Module 4: Prompting and Retrieval-Augmented Generation (RAG)

Week 15-16: Begin Retrieval-Augmented Generation

- Learn what Retrieval-Augmented Generation (RAG) is.
- Study LangChain's indexes and retrievers.
- Understand data ingestion methods.
- Explore text splitters and their usefulness.
- Build a simple customer support Q&A chatbot.
- Learn about embeddings.
- Understand LangChain Chains.
- Build a YouTube video summarizer.
- Explore how to build a voice assistant for your knowledge base.
- Study techniques for preventing undesirable outputs.

Module 5: Advanced RAG and Agents

Advanced RAG and Agents

Explore more advanced RAG techniques and understand how to use LLMs as agents.

Week 17-18: Focus on Advanced RAG

- Understand the differences between prompting, fine-tuning, and RAG.
- Explore advanced RAG techniques with Llama Index.
- Learn about production-ready RAG solutions with LlamaIndex.
- Study RAG metrics and evaluation.
- Learn about LangChain's LangSmith.

Module 5: Advanced RAG and Agents

Week 19-20: Focus on Agents

- Understand what agents are and how large models are used as reasoning engines.
- Explore AutoGPT and BabyAGI.
- Study agent simulation projects in LangChain.
- Build an agent for analysis report creation.
- Learn how to query and summarize a database with LlamaIndex.
- Explore building agents with OpenAI Assistants.
- Learn about LangChain OpenGPT.
- Work on a project for multimodal financial document analysis from PDFs.

Module 6: Fine-Tuning and Deployment

Fine-Tuning and Deployment

Understand how to fine-tune LLMs and deploy them effectively.

Week 21-22: Learn about Fine-Tuning

- Study techniques for fine-tuning LLMs.
- Learn about Low-Rank Adaptation (LoRA).
- Work on a practical example using SFT with LoRA.
- Study using SFT for financial sentiment.
- Learn how to fine-tune a Cohere LLM with medical data.
- Explore reinforcement learning from human feedback (RLHF).
- Work on improving LLMs with RLHF.

Module 6: Fine-Tuning and Deployment

Week 23-24: Deployment Considerations

- Explore the challenges of LLM deployment.
- Learn about model quantization.
- Study model pruning.
- Explore deploying an LLM on a cloud CPU.