A Comprehensive Guide for Large Language Models

Duration: 40 hours

Overview

Unlock the power of Large Language Models (LLMs) in this comprehensive course designed for AI enthusiasts and professionals. From foundational knowledge to advanced deployment techniques, this course offers a step-by-step guide through the intricacies of LLMs, ethical considerations, and real-world applications.

Important Note: All demonstrations and labs are performed on the Free Tier of Google Colab (T4/P100 GPU). For larger models (7B+ parameters) or longer training, Colab Pro or local hardware with 24GB+ VRAM is recommended.

Audience Profile

- Data Scientists and Machine Learning Engineers:
- AI Researchers and Academics
- Software Developers:
- IT Professionals
- Business Analysts and Decision Makers
- Students and Enthusiasts

Course Syllabus

Module 1: Introduction to LLMs and Setup

- What are Large Language Models (LLMs)?
- Overview of Open Source LLMs: DeepSeek, Llama, GPT-J
- Local Deployment vs Cloud Deployment

- Hardware Requirements & Options
- Setting Up Development Environment (Colab Free Tier, Local Machine optional)
- Lab - Set Up Google Colab Environment for LLMs

Module 2: Understanding and Adapting LLMs

- Anatomy of pre-trained LLMs
- Tokenization and Embedding Mechanisms
- Fine-tuning Methods (Full, LoRA, QLoRA)
- Best Practices for Sensitive Data Fine-tuning
- Lab - Fine-tune DeepSeek Coder 6.7B (Quantized) on a Small Dataset

Module 3: Data Preparation & Responsible AI

- Collecting and Preparing Sensitive Data
- Data Cleaning & Tokenization
- Anonymization & Privacy
- Responsible AI & Ethical Considerations
- Lab - Preprocess and Prepare Sensitive Data for Fine-tuning

Module 4: Deployment and Optimization

- Exporting and Running Models Locally
- Llama.cpp / GGUF Deployment (CPU-based)
- Model Optimization for Inference
- API Deployment (FastAPI / Gradio / Streamlit)

Module 5: Scaling, Security & Real-world Applications

- Scaling (Colab Pro, Local GPU, Cloud GPUs)
- Security Best Practices
- Real-world Use Cases: Healthcare, Finance, Legal
- Q&A / Wrap-up
- Lab - Deploy a Secure AI Agent Powered by DeepSeek Model