Course Syllabus: Generative AI Developer

Course Title: Generative AI Development: From Foundation Models to Production

Target Audience: This course is for developers, data scientists, and engineers who have a solid understanding of programming (specifically Python) and are looking to specialize in building applications with generative AI.

Course Level: Advanced Intermediate to Expert.

Duration: 12 Weeks

Course Description: This course provides a comprehensive, hands-on path to becoming a Generative AI Developer. The curriculum focuses on the practical skills needed to design, develop, and deploy applications using generative models like LLMs and diffusion models. You will learn how to integrate these powerful models into software, optimize their performance, and manage them in production environments. By the end of this course, you will have the skills and a portfolio of projects to build real-world, innovative AI solutions.

Learning Objectives

Upon successful completion of this course, students will be able to:

- Understand the core architectures of generative models (e.g., Transformers, GANs, VAEs).
- Master the use of major AI frameworks and libraries like Hugging Face, PyTorch, and TensorFlow.
- Develop and fine-tune generative models for specific tasks.
- Design and build full-stack applications that integrate generative AI APIs.
- Apply MLOps principles to deploy, monitor, and scale generative models in the cloud.
- Implement advanced techniques like Retrieval-Augmented Generation (RAG) to enhance model accuracy.
- Recognize and mitigate the ethical and safety risks associated with generative AI.

Course Structure: A Step-by-Step Learning Path

Part 1: Generative Models & Core Frameworks (Weeks 1-4)

This section builds a deep understanding of the generative models that power AI applications and the frameworks used to build them.

Week 1: Fundamentals of Generative AI

- The evolution of generative models: from GANs and VAEs to Diffusion Models and Transformers.
- Key architectures: Understanding the attention mechanism and its role in modern LLMs.
- The Generative AI ecosystem: an overview of key players and open-source models.
- **Hands-on Lab:** Use a pre-trained model from the Hugging Face Hub for a simple task.

Week 2: The Hugging Face Ecosystem

- Introduction to the Hugging Face `transformers` library for working with LLMs.
- Using the `diffusers` library for image and video generation.
- Exploring the Hugging Face `datasets` and `accelerate` libraries for efficient development.
- **Hands-on Project:** Create a simple text-to-image application using the `diffusers` library.

Week 3: PyTorch & TensorFlow for Generative AI

- A deep dive into PyTorch and TensorFlow: key concepts and differences.
- Building and training a simple generative model from scratch.
- Customizing pre-trained models for new applications.
- **Hands-on Lab:** Train a small generative model on a custom dataset.

Week 4: Model Fine-Tuning

- Why fine-tune a model? The difference between pre-training and fine-tuning.
- Preparing a dataset for fine-tuning.
- Practical techniques for fine-tuning LLMs (e.g., LoRA, QLoRA).
- **Hands-on Project:** Fine-tune an open-source LLM for a specific task, such as generating text in a particular style.

Part 2: Application Development & Advanced Techniques (Weeks 5-8)

This section focuses on integrating models into real-world applications and applying advanced techniques to improve performance and reliability.

Week 5: Building AI APIs & Web Applications

- Designing a RESTful API for your generative model using frameworks like **FastAPI** or **Flask**.
- Building a simple web interface (e.g., using Streamlit or Gradio) to interact with the API.
- **Hands-on Project:** Create a full-stack web application that uses your fine-tuned model to generate content.

Week 6: Advanced Prompting & Agentic AI

- Reviewing advanced prompting techniques (e.g., Chain-of-Thought, persona prompting).
- Introduction to Agentic AI and the concept of agents.
- Building agents with frameworks like **LangChain** or **LlamaIndex** to automate multi-step tasks.
- **Hands-on Lab:** Build an AI agent that can perform a series of actions, such as researching a topic and writing a summary.

Week 7: Retrieval-Augmented Generation (RAG)

- The RAG paradigm: How to ground an LLM with your own data to reduce "hallucinations."
- Understanding vector embeddings and vector databases.
- Building a RAG pipeline from scratch to enable a chatbot to answer questions about a private document.
- **Hands-on Project:** Create a chatbot that can answer questions about a PDF document using RAG.

Week 8: Multimodal & Cross-Generative Tasks

- Working with multimodal models (e.g., Google's Gemini, OpenAI's GPT-4V).
- Combining different generative tasks (e.g., text-to-image, image-to-text, text-to-audio).
- **Hands-on Project:** Build a tool that can analyze an image and then generate a short story based on its contents.

Part 3: MLOps, Cloud Deployment & Expert Skills (Weeks 9-12)

This final section focuses on taking an AI application from development to a professional, scalable production environment.

Week 9: MLOps for Generative AI

- The MLOps lifecycle: data ingestion, model training, serving, and monitoring.
- Containerization with **Docker** for consistent and reproducible deployments.
- Version control for models and datasets.
- **Hands-on Lab:** Dockerize your generative AI application for easy deployment.

Week 10: Cloud Deployment & Scaling

- Deploying a generative AI application on a major cloud platform (e.g., AWS, GCP, Azure).
- Using managed services for model serving and inference.
- Strategies for scaling generative AI applications to handle high traffic.
- **Hands-on Project:** Deploy your Dockerized RAG chatbot to the cloud.

Week 11: Responsible AI & Ethics in Generative Systems

- Identifying and mitigating bias and toxicity in generative models.
- Implementing safety guardrails and moderation APIs.
- The legal and ethical implications of using AI-generated content.
- **Hands-on Lab:** Implement a moderation layer to filter potentially harmful content generated by your model.

Week 12: Final Capstone Project & Career Skills

- **Capstone Project:** Design, build, and deploy a complete, professional-grade generative AI application from scratch. This could be a personalized content generator, a code-assistant, or a creative design tool.
- Building a professional portfolio and resume tailored for Generative AI Developer roles.
- Interview preparation and understanding the current industry landscape.

Assignments & Grading

- **Weekly Hands-on Labs & Exercises:** 20%
- **Intermediate Projects (Weeks 4, 7, 8):** 30%
- **Final Capstone Project:** 40%
- **Code Quality & Best Practices:** 10%