

LLM: Application through Production Syllabus

Course Description

This course is aimed at developers, data scientists, and engineers looking to build LLM-centric applications with the latest and most popular frameworks. You will use Hugging Face to solve NLP problems, leverage LangChain to perform complex, multi-stage tasks, and deep dive into prompt engineering. You will use data embeddings and vector databases to augment LLM pipelines. Additionally, you will fine-tune LLMs with domain-specific data to improve performance and cost, as well as identify the benefits and drawbacks of proprietary models. You will assess the societal, safety, and ethical considerations of using LLMs. Finally, you will learn how to deploy your models at scale, leveraging LLMOps best practices.

By the end of this course, you will have built an end-to-end LLM workflow that is ready for production!

Learning Outcomes

After completing this course, you will be able to:

- Apply LLMs to real-world problems in natural language processing using popular libraries, such as Hugging Face and LangChain.
- Build a custom chat model leveraging open-source LLMs
- Understand the theory behind foundation models, how to fine-tune foundation models on custom datasets, and the innovations that led to GPT-4 and ChatGPT.
- Implement LLMOps and multi-step reasoning best practices.
- Evaluate the efficacy and bias of LLMs using different methods.

Course Content and Activities

Module	Lessons	Assignments (Verified ONLY)
Introduction	Introduction by Matei Zaharia Primer on NLP	Quiz (ungraded)

	<p>Language Models Tokenization Word Embeddings Summary Notebook Demo</p>	
1 - Applications with LLMs	<p>Module Overview Hugging Face Model Selection NLP Tasks Prompts Prompt Engineering Summary Notebook Demo</p>	<p>Quiz #1 Lab #1</p>
2 - Embeddings, Vector Databases, and Search	<p>Module Overview How does vector search work? Filtering Vector Stores Best Practices Summary Notebook Demo</p>	<p>Quiz #2 Lab #2</p>
3 - Multi-stage Reasoning	<p>Module Overview Prompt Engineering LLM Chains Agents Summary Notebook Demo</p>	<p>Quiz #3 Lab #3</p>
4 - Fine-tuning and Evaluating LLMs	<p>Module Overview Applying Foundation LLMs Fine-Tuning: Few-shot learning Fine-Tuning: Instruction-following LLMs Fine-Tuning: LLMs-as-a-service Fine-Tuning: DIY Evaluating LLMs Task-specific Evaluation Summary Notebook Demo</p>	<p>Quiz #4 Lab #4</p>
5 - Society and LLMs	<p>Module Overview Risks and Limitations Hallucination Mitigation Strategies</p>	<p>Quiz #5 Lab #5</p>

	Summary Notebook Demo	
6 - LLMOps	Module Overview Traditional MLOps LLMOps LLMOps Details Summary Notebook Demo	Quiz #6

Prerequisites

- Intermediate-level experience with Python
- Working knowledge of machine learning and deep learning is helpful

Grading

- 40% Quizzes (6 graded in total)
- 60% Labs (5 in total)

Estimated Effort

- 4-12 hours/week, 6 weeks total

Languages

Content: English | Videos: English | Transcripts: English

Enrollment Tracks

- Audit - Freely experience the course during the preview period.
- Verified - Receive a verified certificate by passing the course with a final grade at or above 70%.
- Cost: \$99 (US)

Accessibility Support

- Visit [edX Website Accessibility Policy](#).
- Contact accessibility@edx.org for accessibility questions, concerns, or feedback.