



# Building with Large Language Models

## Leveraging LLMs in the Age of Generative AI

### Your Learning Transformation

By the end of this course, you will be able to **build** and **deploy** complex AI applications that produce **high-quality** outputs using the latest in open-source LLM technology.

### Learning Outcomes

By the end of this workshop you will:

1. Learn the essentials of prompting and fine-tuning
2. Fine-tune the structure of open-source LLMs using a data-centric approach
3. Create a ChatGPT-like interface for your own data using indexing and chaining
4. Build and deploy three different Generative AI applications

### Project-Based Learning

In this workshop, you will deploy three applications in three weeks, including your own unique LLM capstone project in Week 3. You'll leverage the most powerful LLMs to build and deploy personalized applications in your domain and for your use cases. You'll also fine-tune LLMs to improve classic ML algorithm performance, teach LLMs new structure, and build a question-answering tool for your documents.

### Technical Requirements and Prerequisites

This workshop is intended for software engineers, data scientists, ML engineers, or others who have strong Python skills and have working experience deploying applications. Before the workshop, you should:

- Create an account on Hugging Face and be familiar with Hugging Face Datasets, Models, and Spaces. We recommend checking out our [Zero to Deployment Adventure Guide](#) from our recent Building Generative AI Applications Workshop, and checking out the [Hugging Face Course](#) for more information.
- Create a [Google Colab Pro](#) account, which is what we will be using for demos during the workshop.



## Detailed Schedule

Module	Topics	Build Activities
<b>Week 1</b> Fine-Tuning and Deploying LLM Applications	<b>Introduction and Networking</b> <ul style="list-style-type: none"> <li>• Program Overview</li> <li>• 2023 LLM Model Update</li> <li>• 2023 LLM Ops Tool Stack Overview</li> <li>• Most Popular Applications</li> </ul> <b>Model Fine-Tuning</b> <ul style="list-style-type: none"> <li>• Tuning Structure vs. Knowledge</li> <li>• Base Models vs. Instruction-Tuned</li> <li>• Hardware, Configuration, Cost, and Latency Considerations</li> <li>• Tuning via APIs vs. Cloud Hardware</li> </ul> <b>Prompting/Prompt Engineering</b> <ul style="list-style-type: none"> <li>• Prompting Best Practices for LLMs</li> <li>• Seeding the Prompt with Fine-Tuning</li> </ul> <b>Special Topics</b> <ul style="list-style-type: none"> <li>• Prompt-Tuning</li> <li>• 1-Click AI Application Deployments</li> <li>• Creating Synthetic Training Data with ChatGPT-4</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Keeping up with LLMs!</b></li> <li>• <b>Fine-Tuning a Simple Classifier</b></li> <li>• <b>Initial Week 3 Project Ideation Activities</b></li> <li>• Interactive End-to-End BLOOMZ Data Curation, Fine-Tuning, and Deployment Demo on Hugging Face</li> <li>• <b><u>Homework: Fine-Tuning BLOOMZ with Google Colab</u></b></li> </ul>
<b>Week 2</b> Indexing & Document Question Answering	<b>Technologies for “Building ChatGPT For Your Data”</b> <ul style="list-style-type: none"> <li>• Indexing, or Structuring Documents for LLMs to Interact with Them</li> <li>• Vector Databases, or Search and Retrieval using Vector Embeddings</li> <li>• Chaining, or Building Complex AI Applications with LLMs</li> <li>• Emerging Tools on the Market (e.g., LangChain, LlamaIndex, HayStack)</li> </ul> <b>Applications Considerations</b> <ul style="list-style-type: none"> <li>• From One Document to Many</li> <li>• From Qualitative Question-Answering to Quantitative</li> <li>• Chatbots versus Virtual Assistants</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Document Data Collection</b></li> <li>• <b>Answering Questions from a Single Document</b></li> <li>• <b>Week 3 Project Commitments</b></li> <li>• Interactive End-to-End Document Question Answering Deployment Demo on Hugging Face</li> <li>• <b><u>Homework: Indexing and Querying Multiple Documents using Indexing</u></b></li> </ul>



	<p><b>Special Topics</b></p> <ul style="list-style-type: none"> <li>Using GPT-4 in Your Development Process</li> <li>The Space of Generative AI Startups</li> </ul>	<p><b>with LangChain and Open AI's GPT 3.5</b></p>
<p><b>Week 3</b> Validating LLM Application Outputs and Demo Day!</p>	<p><b>How Do I Know If My Output is Good?</b></p> <ul style="list-style-type: none"> <li>The Importance of Assessing Quality for Your Use Case</li> <li>Emerging Metrics for Assessing LLM Outputs (e.g., Perplexity, Burstiness, etc.)</li> <li>LLM Model Cards of the Future</li> <li>LLM Ops: When to Trigger Retraining of a Production LLM</li> </ul> <p><b>Emerging Benchmarks &amp; Best-Practices</b></p> <ul style="list-style-type: none"> <li>Choosing the Right LLM for your Job (Stanford's Holistic Evaluation of Large Language Models)</li> <li>LLM Trade Offs: Size and Capability vs. Cost</li> <li>Legal Considerations</li> </ul> <p><b>Special Topics</b></p> <ul style="list-style-type: none"> <li>Building Your Generative AI Brand</li> <li>An Outlook on Jobs in 2023</li> </ul>	<ul style="list-style-type: none"> <li>Interactive Demo of Proposed Validation Framework and Model Card for Structure-Fine-Tuned BLOOMZ Application from Week 1</li> <li><b>Discussion of Potential Ways to Improve Quality Validation Framework</b></li> <li><b>PROJECT: Curate Data, Build a Structure-Fine-Tuned AI Application. Deploy on Hugging Face and Develop a Quality Framework to Include in Your Model Card</b></li> <li><b>Create Final Project Demo to share</b></li> </ul>

\* Activities in **Bold Pink** will be working sessions during the class!

[Register for Building with LLMs](#)

### About FourthBrain

FourthBrain trains engineers, developers, data scientists, and leaders to make an impact in the Artificial Intelligence field, with our flexible, accessible education programs. We are training a new generation of engineers and leaders who have more than just technical ability; they have an awareness and mindset of what is needed to succeed with AI. We are part of the AI Fund, founded by Andrew Ng.