



AI Fine Tuning and Data Prep

- 3-Day Class
- Hands-on labs

Course Overview

You will develop the skills to gather, clean, and organize data for fine-tuning pre-trained LLMs and Generative AI models. Through a combination of lectures and hands-on labs, you will use Python to fine-tune open-source Transformer models. Gain practical experience with LLM frameworks, learn essential training techniques, and explore advanced topics such as quantization. During the hands-on labs, you will access a GPU-accelerated server for practical experience with industry-standard tools and frameworks.

Review this course online at <https://www.alta3.com/courses/00-orphans>

Who Should Attend

- Project Managers
- Architects
- Developers
- Data Acquisition Specialists

What You'll Learn








- Clean and Curate Data for AI Fine-Tuning
- Establish guidelines for obtaining RAW Data
- Go from Drowning in Data to Clean Data
- Fine-Tune AI Models with PyTorch
- Understand AI architecture: Transformer model
- Describe tokenization and word embeddings
- Install and use AI frameworks like Llama-3
- Perform LoRA and QLoRA Fine-Tuning
- Explore model quantization and fine-tuning
- Deploy and Maximize AI Model Performance

Outline


Data Curation for AI

- 🗨️ Lecture: Curating Data for AI
- 🖥️ Lecture + Lab: Gathering Raw Data
- 🖥️ Lecture + Lab: Data Cleaning and Preparation
- 🖥️ Lecture + Lab: Data Labeling
- 🖥️ Lecture + Lab: Data Organization
- 🗨️ Lecture: Premade Datasets for Fine Tuning
- 🖥️ Lecture + Lab: Obtain and Prepare Premade Datasets






Deep Learning

-  Lecture: What is Intelligence?
-  Lecture: Generative AI
-  Lecture: The Transformer Model
-  Lecture: Feed Forward Neural Networks
-  Lecture + Lab: Tokenization
-  Lecture + Lab: Word Embeddings
-  Lecture + Lab: Positional Encoding





Pre-trained LLM

-  Lecture: A History of Neural Network Architectures
-  Lecture: Introduction to the LLaMa.cpp Interface
-  Lecture: Preparing A100 for Server Operations
-  Lecture + Lab: Operate LLaMa3 Models with LLaMa.cpp
-  Lecture + Lab: Selecting Quantization Level to Meet Performance and Perplexity Requirements

Fine Tuning

-  Lecture: Fine-Tuning a Pre-Trained LLM
-  Lecture: PyTorch
-  Lecture + Lab: Basic Fine Tuning with PyTorch
-  Lecture + Lab: LoRA Fine-Tuning LLaMa3 8B
-  Lecture + Lab: QLoRA Fine-Tuning LLaMa3 8B

Operating Fine-Tuned Model

-  Lecture: Running the llama.cpp Package
-  Lecture + Lab: Deploy Llama API Server
-  Lecture + Lab: Develop LLaMa Client Application
-  Lecture + Lab: Write a Real-World AI Application using the Llama API

Prerequisites

- Python - PCEP Certification or Equivalent Experience
- Familiarity with Linux

Next Courses

Course 1 Course 2