



Open-Source Generative AI

COURSE OVERVIEW

Open-Source Generative AI for the Enterprise is a comprehensive 4-day course that teaches practical applications for AI in the business environment. This course offers a combination of lectures and hands-on labs, providing participants with a solid understanding of AI concepts and the skills to design and implement AI solutions.

Throughout the course, you will learn about AI transformer-based architectures, the fundamentals of Python programming for AI deployments, and the deployment of open-source Transformer models. You will also explore the importance of hardware requirements in AI performance, comparing different GPU architectures and understanding how to match AI requirements with suitable hardware. The course delves into training techniques, including backpropagation, gradient descent, and various AI tasks such as classification, regression, and clustering.

You will gain practical experience through hands-on exercises with open-source LLM (Language Learning Model) frameworks, allowing you to work with fine-tuned models and run workloads on different models to understand their strengths and weaknesses. Additionally, the course covers the conversion of model formats and provides an in-depth exploration of AI programming environments like PyTorch + transformers and transformers' low-level interactive inspection.

Towards the end of the course, you will delve into advanced topics such as context extension through fine-tuning and quantization for specific application target environments. By the completion of the course, you will have the opportunity to earn an AI certification from Alta3 Research, further enhancing your credentials in the field of Artificial Intelligence. This course is ideal for Python Developers, DevSecOps Engineers, and Managers or Directors seeking a practical overview of AI and its practical application in the enterprise.

WHO WILL BENEFIT FROM THIS COURSE?

- Project Managers
- Architects
- CKA Developers
- Data Acquisition Specialist

PREREQUISITES

Previous exposure to any programming language, preferably Python

COURSE OBJECTIVES

- 5G EN-DC architecture
- 5G Standalone architecture
- Beamforming, mm-wave, massive-mimo
- 5G Access technology, O-RAN, and v-RAN



- New Radio technology, numerology
- Narrowband IoT support technology
- IMS in the 5G Network
- Network Slicing
- 5G call trace analysis

COURSE OUTLINE

The Mechanics of Deep Learning - Gain an intuitive understanding of the most current generative AI architecture.

- Choosing the latest AI models
 - What is current and what has already gone extinct?
- Neural network architecture essentials
 - Tokenization
 - Embedding
 - Parameters: weights and bias
 - Nodes
 - Fully connected/Partially connected
 - Prompts and prompt engineering

The Transformer Model - Develop an intuitive understanding of the transformer model, without math.

- Neural Network Architectures
- Word embeddings
 - Importance in NLP
 - Word2Vec and GloVe embeddings
 - Contextualized embeddings (BERT, ELMo, etc.)
- Self-attention mechanism and multi-head attention
 - Input Representation (Query, Key, and Value)
 - Computing Similarities
 - Attention Weights
 - Weighted Sum
 - Final Output
- Positional encoding for attention
- Transformer layers and stacking
- Quantization
- Effective LLM selection criteria and use cases
 - Models: size, datasets, quantization, etc
 - Libraries
 - Frameworks

Hardware Requirements

- GPUs role in AI performance (CPU vs GPU)
- GPUs architecture history
 - Pascal, Turing, Ampere
- Tensor core vs older GPU architectures
- GPU vocabulary and the transformer



- SM, tensors, tensor core, CUDA core, FP and INT cores, warp, threads
- Current GPUs and cost vs value
 - Analysis of GPU specification sheets
- GPU selection for models and workload
- Quantizing for hardware performance and cost

Pre-trained LLM Model Essentials

- Select and use pre-trained models to produce immediate results
- Model ratings, metrics, leader boards
- Model licensing and commercial use
- Numbers every AI developer should know (cost, time, dataset size, etc.)
- Synthetic data generation for model training.
- Model training through feature extraction
 - Derive features from the provided dataset
 - Use annotated data for fine-tuning training

Pre-trained LLM hands-on - Run inference on multiple models on diverse prompts

- Frameworks
 - llama.cpp, exllama, GPT4All
- Evaluate multiple models, fine-tuning quantization
 - Falcon, Orca Mini, OpenLlama, Alpaca, MPT
- Parameters
- Fine-tuned models
- Prompts
 - Model expectations (instruct, chat, etc)
 - Extension improvement

Transformer Model training essentials

- Overfitting
- Regularization (Avoiding Overfitting)
- Backpropagation
- Gradient Descent
- Embedding
- Learning Rate
- Perplexity
- Batch Normalization
- Warm-up and Learning Rate Decay
- Loss Functions
- Data Augmentation
- Training Strategies
- Evaluation Metrics

Conversion of Model Formats

- PyTorch to ggml
- JAX to ggml
- F16 to 4bit quantization



Hands-on with AI programming environments

- PyTorch and Transformers
- Transformers low-level interactive inspection

Model Fine Tuning

- Perform fine-tuning in a hands-on environment
- Demonstration project with good fine-tuning data to highlight frameworks
 - with llama.cpp
 - with PyTorch
- Understanding fine-tuning dataset formats
- Data cleaning skills

Introduce Application Interfacing to Models

- LangChain
- Guidance for structured LLM output

Applications Augmentation with Langchain and Guidance

- LangChain
- Guidance for structured LLM output
- Prompt engineering

Advanced Topics

- Optimize your model through fine-tuning and quantization.
- Realize quality and speed inference on the specific workload
- Context extension through fine-tuning
- LLM deployment

Use llama to perform Natural Language Processing tasks

- Rewrite a classic poem of a well-known author in the tone of another well-known author
- Use Named Entity Recognition (NER) to identify Cajun food in recipes
- Use domain-specific models to improve project accuracy

Deploy a Natural Language Model Capstone

- Build a full NLP project
- Download, install, and implement a trained NLP model
- Deploy your NLP to perform the following tasks:
 - text classification
 - sentiment analysis
 - machine translation
 - chatbots
 - question answering

WHY TRAIN WITH SUNSET LEARNING INSTITUTE?

Sunset Learning Institute (SLI) has been an innovative leader in developing and delivering authorized technical training since 1996. Our goal is to help our customers optimize their technology Investments by providing convenient, high quality technical training that our customers can rely on. We empower students to master their desired technologies for their unique environments.

What sets SLI apart is not only our immense selection of trainings options, but our convenient and consistent delivery system. No matter how complex your environment is or where you are located, SLI is sure to have a training solution that you can count on!

Premiere World Class Instruction Team

- All SLI instructors have a four-year technical degree, instructor level certifications and field consulting work experience
- Sunset Learning has won numerous Instructor Excellence and Instructor Quality Distinction awards since 2012

Enhanced Learning Experience

- The goal of our instructors during class is ensure students understand the material, guide them through our labs and encourage questions and interactive discussions.

Convenient and Reliable Training Experience

- You have the option to attend classes live with the instructor, at any of our established training facilities, or from the convenience of your home or office
- All Sunset Learning Institute classes are guaranteed to run – you can count on us to deliver the training you need when you need it!

Outstanding Customer Service

- You will work with a dedicated account manager to suggest the optimal learning path for you and/or your team
- An enthusiastic student services team is available to answer any questions and ensure a quality training experience

Interested in Private Group Training?

[Contact Us](#)