Training Course on Large Language Models (LLMs) from Scratch

Professional Development Training Course Outline

Category	Data Science	**Duration**	10 Days
Base Fee	$3,000 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

Training Course on Large Language Models (LLMs) from Scratch: Understanding Transformer Architecture and Pre-training provides a comprehensive deep dive into the foundational principles of Large Language Models (LLMs), focusing specifically on building them from the ground up. Participants will gain a robust understanding of the revolutionary **Transformer architecture** and the critical **pre-training** phase, equipping them with the theoretical knowledge and practical skills necessary to design, implement, and optimize custom LLMs. This program addresses the growing demand for **AI specialists** capable of developing **generative AI solutions** and leveraging **natural language processing (NLP)** at an advanced level.

Through a blend of theoretical instruction, hands-on coding exercises, and real-world **case studies**, this course demystifies the complexities of LLMs, from **attention mechanisms** and **tokenization** to large-scale data ingestion and **distributed training**. We will explore the nuances of various **pre-training objectives** and delve into the computational challenges and strategies for efficient model development. This course is engineered to empower professionals to contribute to the cutting-edge of **AI innovation**, enabling them to harness the power of **transformative AI** for diverse applications.

Programme Curriculum

Training Course on Large Language Models (LLMs) from Scratch: Understanding Transformer Architecture and Pre-training

Introduction

Training Course on Large Language Models (LLMs) from Scratch: Understanding Transformer Architecture and Pre-training provides a comprehensive deep dive into the foundational principles of Large Language Models (LLMs), focusing specifically on building them from the ground up.

Participants will gain a robust understanding of the revolutionary **Transformer architecture** and the critical **pre-training** phase, equipping them with the theoretical knowledge and practical skills necessary to design, implement, and optimize custom LLMs. This program addresses the growing demand for **AI specialists** capable of developing **generative AI solutions** and leveraging **natural language processing (NLP)** at an advanced level.

Through a blend of theoretical instruction, hands-on coding exercises, and real-world **case studies**, this course demystifies the complexities of LLMs, from **attention mechanisms** and **tokenization** to large-scale data ingestion and **distributed training**. We will explore the nuances of various **pre-training objectives** and delve into the computational challenges and strategies for efficient model development. This course is engineered to empower professionals to contribute to the cutting-edge of **AI innovation**, enabling them to harness the power of **transformative AI** for diverse applications.

Course Duration

10 days

Course Objectives

1. Gain a profound understanding of the core components and mathematical underpinnings of the **Transformer neural network**.
2. Comprehend **self-attention** and multi-head attention, crucial for contextual understanding in LLMs.
3. Understand how Transformers capture sequential information without recurrence.
4. Learn various **tokenization strategies** and their impact on model performance.
5. Grasp the concept of **vector representations** for words and their generation.
6. Develop practical skills in coding and assembling Transformer encoder and decoder blocks.
7. Understand the objectives and methodologies behind **unsupervised pre-training** of LLMs.
8. Learn best practices for collecting, cleaning, and preparing **massive text datasets** for pre-training.
9. Identify and strategize for the significant **compute resources** and **distributed training** required for LLMs.
10. Learn to assess the performance of pre-trained LLMs using relevant metrics and benchmarks.
11. Gain insights into modern Transformer variants and their optimizations (**e.g., GPT, BERT, T5**).
12. Establish a strong base for future **LLM fine-tuning** and **transfer learning** applications.
13. Discuss the ethical implications and biases inherent in **large-scale data training**.

Organizational Benefits

- Equips teams with the expertise to integrate and develop **cutting-edge LLM solutions** within the organization.
- Fosters in-house capability for **generative AI research** and development, leading to novel products and services.
- Reduces reliance on external vendors by enabling internal development and optimization of **custom LLMs**.
- Deepens organizational understanding of **natural language data** and its potential for business insights.
- Positions the organization at the forefront of **AI technology**, leading to **disruptive innovation** and market leadership.
- Facilitates the development of **AI-powered automation** for various text-based tasks.

- Enables organizations to build and train LLMs on their own data, maintaining full control over sensitive information.

Target Audience

1. Machine Learning Engineers
2. Data Scientists
3. AI Researchers
4. Software Developers
5. PhD Students and Academics focusing on advanced AI and computational linguistics.
6. Technical Leads and Architects evaluating LLM integration strategies.
7. Professionals with a strong Python programming background and foundational machine learning knowledge.
8. Anyone eager to build Large Language Models from Scratch.

Course Outline

Module 1: Introduction to Large Language Models & Their Evolution

- What are LLMs? Definition and significance in modern AI.
- Historical context: From traditional NLP to deep learning.
- The rise of **Generative AI** and its impact.
- Overview of popular LLMs (e.g., GPT series, BERT, Llama).
- Case Study: Analyzing the paradigm shift brought by the GPT-3 architecture.

Module 2: Fundamentals of Neural Networks for NLP

- Recap of Artificial Neural Networks (ANNs) and Deep Neural Networks (DNNs).
- Introduction to Recurrent Neural Networks (RNNs) and LSTMs: Their strengths and limitations.
- The vanishing/exploding gradient problem in sequential models.
- Introduction to the need for **Transformer architecture**.
- Case Study: Comparing performance of RNNs vs. early Transformers on sequence tasks.

Module 3: Introduction to Transformer Architecture: "Attention Is All You Need"

- The seminal paper and its core ideas.
- Encoder-Decoder structure overview.
- The concept of **Attention** as a mechanism.
- Parallelization benefits over recurrent models.
- Case Study: Deconstructing the original Transformer paper for key insights.

Module 4: Deep Dive into Self-Attention

- Query, Key, and Value vectors.
- Calculating attention scores and softmax.
- Scaled Dot-Product Attention explained.
- Intuition behind self-attention: how models focus on relevant parts of input.
- Case Study: Visualizing attention weights in a simple sentence completion task.

Module 5: Multi-Head Attention Mechanism

- Why multiple "heads"? Benefits of diverse attention subspaces.
- Concatenation and linear projection of attention outputs.

- Combining information from different attention perspectives.
- Understanding the increased representational power.
- Case Study: Analyzing how different attention heads capture various linguistic relationships.

Module 6: Positional Encodings

- The challenge of sequential information in parallel processing.
- Absolute vs. Relative Positional Encodings.
- Mathematical formulation of sinusoidal positional encodings.
- Integrating positional information into embeddings.
- Case Study: Impact of different positional encoding schemes on translation quality.

Module 7: Feed-Forward Networks and Layer Normalization

- Role of position-wise fully connected feed-forward networks.
- Activation functions (e.g., GELU, ReLU) in Transformers.
- Significance of **Layer Normalization** and Residual Connections.
- Ensuring stable training and better gradient flow.
- Case Study: Observing the effect of normalization on model convergence and performance.

Module 8: Tokenization Strategies for LLMs

- Character, Word, and Subword Tokenization.
- Byte-Pair Encoding (BPE) and WordPiece algorithms.
- SentencePiece and Unigram models.
- Vocabulary size and its implications.
- Case Study: Comparing tokenization schemes for a specific language and their impact on model size and rare word handling.

Module 9: Building the Transformer Encoder

- Putting together Self-Attention, Feed-Forward, and Normalization layers.
- Stacking multiple encoder blocks.
- Input embeddings and positional encodings in the encoder.
- Understanding the encoder's role in contextual representation.
- Case Study: Implementing a simplified Transformer encoder for a sentiment analysis task.

Module 10: Building the Transformer Decoder

- Masked Multi-Head Self-Attention for autoregressive generation.
- Encoder-Decoder Attention (Cross-Attention).
- Stacking decoder blocks and the output layer.
- Generating text token by token.
- Case Study: Building a simple text generation model using a Transformer decoder.

Module 11: Pre-training Objectives for LLMs

- **Language Modeling** (Causal LM) - predicting the next token.
- Masked Language Modeling (MLM) - BERT's approach.
- Next Sentence Prediction (NSP) and other auxiliary tasks.
- The concept of **self-supervised learning**.
- Case Study: Examining the training process of BERT and GPT-2 to understand their distinct pre-training objectives.

Module 12: Data Preparation and Scaling for Pre-training

- Curating **massive text datasets** (e.g., Common Crawl, Wikipedia, books).
- Data cleaning, deduplication, and filtering techniques.
- Strategies for handling noisy and diverse data.
- Importance of data quality for robust LLMs.
- Case Study: A practical walkthrough of preparing a large corpus for pre-training, addressing common data challenges.

Module 13: Computational Aspects of Large-Scale Pre-training

- Hardware requirements: GPUs, TPUs, and **distributed computing**.
- Parallelization strategies: Data Parallelism, Model Parallelism, Pipeline Parallelism.
- Memory optimization techniques (e.g., mixed-precision training, gradient checkpointing).
- Estimating computational costs and timeframes.
- Case Study: Simulating a distributed training setup and analyzing resource utilization.

Module 14: Training and Evaluation of a Base LLM

- Setting up a training pipeline with a deep learning framework (e.g., PyTorch, TensorFlow).
- Optimizer selection (e.g., AdamW) and learning rate schedules.
- Monitoring training progress and identifying common issues.
- Evaluation metrics for generative models (e.g., Perplexity, BLEU, ROUGE).
- Case Study: Training a small Transformer model from scratch on a public dataset and evaluating its performance.

Module 15: Beyond Basic Pre-training & Future Trends

- Introduction to advanced pre-training techniques (e.g., contrastive learning, denoising autoencoders).
- Current research directions in **LLM scalability** and efficiency.
- The role of open-source models and the **Hugging Face Transformers library**.
- Ethical considerations, bias mitigation, and responsible AI development.
- Case Study: Discussing the challenges and future directions in creating truly **interpretable and steerable LLMs**.

Training Methodology

This course employs a highly interactive and practical training methodology, combining:

- **Lectures & Presentations**: Clear explanations of complex concepts, enhanced with visual aids.
- **Hands-on Coding Labs**: Practical exercises using Python and popular deep learning frameworks (PyTorch/TensorFlow) to build and train Transformer components and small LLMs.
- **Code Walkthroughs & Demos**: Step-by-step guidance through practical implementations.
- **Case Studies & Discussions**: Analysis of real-world scenarios and existing LLM architectures, fostering critical thinking.
- **Group Activities & Collaboration**: Encouraging peer learning and problem-solving.
- **Q&A Sessions**: Dedicated time for addressing participant queries and fostering deeper understanding.
- **Resource Sharing**: Providing access to relevant research papers, code repositories, and documentation.

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call *+254769199797*

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to Fineskill Training Center account, as indicated in the invoice so as to enable us prepare better for you.

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Online	$2,000
23 Nov 2026	04 Dec 2026	Physical	$3,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$12,000

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Physical	$9,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$12,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com