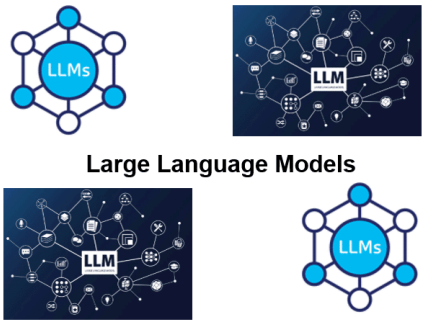
Large Language Models

Duration: 2 days Price: € 1499

Open Schedule



Audience course Large Language Models

The course Large Language Models is intended for software engineers, data scientists, and technical professionals who want to work with large language models (LLMs).

Prerequisites Large Language Models Course

To participate in the course, a basic understanding of Python and machine learning is required. Familiarity with neural networks or natural language processing is useful.

Realization training Large Language Models

The course is led by an experienced trainer and includes a mix of theory and hands-on exercises. Demonstrations and case studies involving LLMs are used to illustrate key concepts.

Large Language Models Certificate

After successfully completing the course, attendants receive a certificate of participation in the course Large Language Models.

Content Large Language Models

The course Large Language Models from SpiralTrain provides a comprehensive understanding of large language models (LLMs), from foundational architectures like transformers to advanced topics like fine-tuning, safety, deployment, and real-world applications. Participants will explore open models, tools, and cutting-edge use cases across domains.

Intro to LLMs

This module introduces LLMs and their evolution, from GPT to BERT and T5. It explains transformers, attention, and tokenization. Participants explore training objectives, scaling laws, and key differences between pretraining and fine-tuning. Open source vs proprietary models are also compared.

Model Architectures

Participants learn LLM types: decoders, encoder-decoders, and key models like GPT, LLaMA, and PaLM. It covers training pipelines, optimizers, and precision formats. Tools like Hugging Face and Deepspeed are introduced, plus tuning techniques like LoRA and in-context learning.

Training LLMs

This module focuses on preparing and fine-tuning data for LLMs. It includes tokenizer setup, adapters, SFT, and avoiding overfitting. Participants learn metrics for evaluation, model alignment, and benchmarking. Hugging Face tools and best practices are emphasized.

LLM Deployment

Participants learn how to serve and optimize LLMs for production. Topics include quantization, distillation, and cloud deployment (AWS, Azure, GCP). Also covered are LangChain integration, embeddings, caching, and reducing inference costs with scalable strategies.

Safety and Bias

Focus is on LLM safety: identifying and reducing bias, model auditing, and prompt attacks. Topics include explainability, red teaming, and moderation. Legal and privacy concerns are discussed, with strategies for responsible LLM deployment.

LLM Use Cases

The course ends with real-world LLM applications in code, legal, health, and education. Use cases include RAG systems, agents, and plugins. Participants explore enterprise integration, model evaluation, and research trends shaping the future of LLMs.

Modules Large Language Models

Module 1: Intro to LLMs	Module 2: Model Architectures	Module 3: Training LLMs
What are LLMs? Transformer architecture Training Objectives (causal, masked) Evolution of LLMs (GPT, BERT, T5) Open Source vs Proprietary LLMs Tokenization and Vocabulary Attention Mechanism Model Scaling Laws Transfer Learning Pretraining vs Fine-Tuning	Decoder vs Encoder-Decoder Models GPT, LLaMA, T5, and PaLM Training Pipeline Overview Optimizers (Adam, Adafactor) Precision (FP32, FP16, quantization) Transformers (HF), Megatron, Deepspeed Parameter vs Instruction Suning LoRA and QLoRA In-context Learning Reinforcement Learning with HF	Dataset Creation and Curation Tokenizer Customization Data Preprocessing Fine-Tuning with Hugging Face SFT (Supervised Fine-Tuning) Adapters and LoRA Evaluation Metrics Avoiding Overfitting Model Alignment Model Evaluation and Benchmarking
Module 4: LLM Deployment	**Module 5: Safety and Bias**	**Module 6: LLM Use Cases**
Inference Optimization Model Distillation Quantization Techniques Hosting on AWS, GCP, Azure Using Model Gateways LangChain and Semantic Search Vector Stores and Embeddings Caching Responses Load Balancing Cost Optimization Strategies	Understanding Model Biases Mitigation Strategies Model Auditing Adversarial Prompts User Privacy Filtering and Moderation Red Teaming Explainability in LLMs Interpreting Outputs Regulatory and Legal Issues	Coding Assistants AI for Legal and Finance Education and Learning Health Care and Biotech Chatbots and Agents RAG Systems Tool Use and Plugins Enterprise Use of LLMs Evaluating New Models Future Directions LLM Research