

AI-110: Intro to Large Language Models (LLMs) and LLM-based apps.

Course Length: 8 training hours

Course Description:

The fast pace of development in LLMs and related technologies made it possible to use them even in enterprise grade applications. There are already a few areas where a new generation of LLM-based applications totally redefined applications' capabilities and users' expectations while AI technologies are going to radically change all kinds of other software areas as well.

That's why software developers as well as other IT professionals and technical managers need to understand these technologies and need to have practical skills to use them in their daily work.

Training objectives: At the end of the training participants:

- Recognize common LLM-based applications and understand their main building blocks.
- Get a high-level understanding of how modern large language models (LLMs) work and how they are trained in multiple steps.
- Explain the benefits of using LLM development frameworks, understand their key features, and be familiar with some popular open-source options.
- Understand the main ideas behind prompt engineering, including practical tips and best practices for working effectively with modern LLMs.
- Know the basics of RAG (Retrieval-Augmented Generation) systems, including their main parts, ways to improve their performance, and some new alternative solutions.
- Understand the motivations for and the two main types of LLM-based agentic systems as well as the key components and the way of working of autonomous agents.
- Recognize the importance of monitoring and evaluating LLM applications throughout their lifecycle as well as learning about some leading tools available and their services.

Main topics:

- Introduction to LLM based applications: current types, building blocks, challenges
- Why and how LLMs work and are trained?
- Using closed- and open-source LLMs directly via their APIs
- Using LLMs via app. development frameworks
- Prompt engineering
- "Talk with your documents": Retrieval Augmented Generation (RAG)
- "AI that thinks and acts": LLM Agents
- Quality Assurance at LLM apps: Tracing and Evaluation (optional)

Besides gaining a basic understanding of Large Language Models (LLMs) and other technologies used in LLM-based applications, students will be able to examine their features and play with them during instructor's demonstration and lab exercises.

This training is part of the AI portfolio of Component Soft which explores essential AI topics, such as:

- AI-110 Intro to Large Language Model (LLMs) and LLM-based apps.
- AI-141: Using Github Copilot as coding assistant
- AI-151: Using Codeium/Windsurf as coding assistant
- AI-161: Using Amazon Q as coding assistant
- AI-434: GenAI Application Development with LLMs
- AI-435: GenAI Application Development with LLMs (extended version)

Structure: 50% lecture, 25% demonstration by the instructor, 25% hands on lab exercises

Target audience: Technical managers as well as IT and telco professionals who want to familiarize themselves with Large Language Models (LLMs) and LLM based applications.

Prerequisites: Basic understanding of IT concepts, User experience with ChatGPT or similar chatbots.

Detailed Course Outline

Module 1. Introduction to LLM based applications: current types, building blocks, challenges

- Main usage areas of LLM-based applications
- Main types of LLM-based applications
- Building blocks of LLM-based applications
- Demo: Popular LLM-based application types

Module 2. Why and how LLMs work and are trained?

- Main elements and operation of LLMs (tokenizer, embeddings, transformer, transformer head, next token selector)
- The 4+1 training phase of LLMs
- Most important LLM vendors and models
- Demo: LLMs before ChatGPT

Module 3. Using closed- and open-source LLMs directly via their APIs

- Using LLMs through APIs
- LLM API text generation modes
- Typical LLM parameters
- Demo: Using a closed-source and an open-source LLM via API

Module 4. Using LLMs via app. development frameworks

- What are LLM app. development frameworks
- Their advantages and disadvantages
- Case study: Langchain, the most popular framework
- Demo: Using LangChain in an LLM chatbot

Module 5. Prompt engineering

- What is prompt engineering?
- Golden rules of prompt engineering
- The “Just Ask” Principle, Zero-shot prompts
- Some important prompt engineering rules
- Demo: Demonstrating basic prompt techniques

Module 6. “Talk with your documents”: Retrieval Augmented Generation (RAG)

- What is Retriever Augmented Generation (RAG)?
- How does RAG work?
- Main building blocks of an RAG pipeline
- Advanced RAG techniques
- **Demo:** Demonstration of Retrieval Augmented Generation (RAG) in an LLM app

Module 7. “AI that thinks and acts”: LLM Agents

- Motivations for LLM-based Agentic Systems
- Main Features of and Difference between Workflows and Agents
- Main Building Blocks: Functions, Tools, Agents
- The ReAct autonomous agent execution logic
- Multi-agent systems
- Demo: Agentic workflow and agent

Module 8. Quality Assurance at LLM apps: Tracing and Evaluation (optional)

- Why do we need them during development?
- Tracing and evaluation tools for LLM-based apps
- Tracing basics
- Evaluation basics
- **Demo:** Langsmith Tracing and Evaluation