

Building Practical AI Systems: From Foundations to Domain-Specialized Local LLMs

Descrizione

Il corso introduce i concetti fondamentali dell'intelligenza artificiale moderna e dei Large Language Models, spiegando in modo chiaro come funzionano i modelli e quali sono le reali possibilità di utilizzo in contesti aziendali.

Partendo dalle basi teoriche dell'AI e dei modelli linguistici, il corso accompagna i partecipanti nella costruzione pratica di un sistema AI completo: dall'esecuzione di modelli locali alla creazione di una knowledge base interrogabile fino alla specializzazione di modelli più piccoli su domini tecnici specifici.

Attraverso laboratori pratici i partecipanti impareranno a:

- eseguire modelli LLM localmente
- trasformare documentazione tecnica in conoscenza interrogabile
- costruire sistemi basati su Retrieval Augmented Generation
- creare dataset specialistici partendo da documentazione tecnica
- adattare modelli più piccoli per domini specifici

Il corso è progettato per fornire una comprensione chiara della tecnologia evitando sia eccessi teorici sia aspettative irrealistiche sul reale utilizzo dell'AI in azienda.

Agenda del corso

1. Introduzione all'Intelligenza Artificiale

Cos'è realmente l'intelligenza artificiale

Breve evoluzione dell'AI e del machine learning

Dal machine learning tradizionale ai foundation models

Terminologia fondamentale: modello, training, inference, dataset

Come "impara" un modello

Differenze tra:

- AI
- Machine Learning
- Deep Learning
- Large Language Models

2. Fondamenti dei Large Language Models

Tokenizzazione

Embedding e rappresentazione semantica

Architettura Transformer

Meccanismo di attenzione

Parametri e parametri attivi

Scaling laws e dimensione dei modelli

Training vs Inference

Parametri di generazione:

- temperature
 - top-p
 - top-k
 - max tokens
-

3. Tipologie di modelli e licenze

Distinzione tra:

Modelli pubblici via API

Modelli open source

Modelli open weight

Modelli installabili localmente

Aspetti legali e licenze

Limitazioni tecniche dei modelli locali

Quando ha senso usare:

- API cloud
 - modelli locali
 - sistemi ibridi
-

Setup dell'ambiente AI locale

4. Ambienti per eseguire LLM in locale

Panoramica degli strumenti per LLM locali

Tecnologie utilizzate nel corso:

Runtime locali

- **Ollama**
- **llama.cpp**

Linguaggio e librerie

- **Python**
- **LangChain / LlamaIndex**

Modelli utilizzati

- **Qwen 3.5 small models**
- modelli open compatibili con runtime locali

Strumenti di supporto

- **Docker** per ambienti isolati
- **VSCode** per sviluppo e testing

Laboratorio 1

Installazione di un LLM locale

Installazione runtime locale

- setup Ollama
- download modello

Test di inferenza

Parametri di generazione

Confronto tra modelli di dimensioni diverse

Knowledge Systems per AI

5. Trasformare documenti in conoscenza

Problemi della documentazione aziendale

Manuali tecnici

PDF

Runbook operativi

Wiki aziendali

Pipeline di trasformazione dei documenti

Tecnologie utilizzate

- **Docling** per parsing documentale
- conversione in **Markdown strutturato**

Preparazione dei contenuti per sistemi AI

Laboratorio 2

Parsing documenti tecnici

Conversione documenti PDF in Markdown

Pulizia e organizzazione dei contenuti

Preparazione dei documenti per indicizzazione

Retrieval Augmented Generation

6. Costruire sistemi AI basati su documentazione

Limiti dei modelli statici

Introduzione al Retrieval Augmented Generation

Embedding e ricerca semantica

Tecnologie utilizzate

- **Sentence Transformers**
- **FAISS / ChromaDB** per vector search

Pipeline RAG

- ingestione documenti
 - indicizzazione
 - retrieval
 - generazione risposta
-

Laboratorio 3

Costruzione di un sistema RAG
Indicizzazione dei documenti
Creazione di una knowledge base interrogabile
Interrogazione tramite LLM locale
Test su manuali tecnici

Teacher Models e dataset

7. Generare dataset specialistici

Uso di modelli avanzati come teacher
Creazione di dataset Q&A
Generazione di esempi sintetici
Valutazione dei dataset

Specializzazione dei modelli

8. Creare modelli specialistici

Differenze tra
fine tuning
distillazione
retrieval
Quando conviene specializzare un modello
Creazione di modelli piccoli per domini specifici
Esempi
LLM per sistemisti
LLM per sviluppatori
LLM per supporto tecnico

Laboratorio 4

Creazione di un dataset specialistico
Adattamento di un piccolo modello
Test di un modello specializzato

Architettura finale

9. Sistemi AI aziendali

Knowledge layer centrale
Modelli specialistici locali
Deployment in azienda
Costi reali
Sicurezza e privacy dei dati

Tecnologie utilizzate nei laboratori

Runtime LLM
Ollama
llama.cpp
Modelli
Qwen 3.5 small models
Pipeline documentale
Docling
Vector database
FAISS / ChromaDB
Framework AI
LangChain / LlamaIndex
Ambiente sviluppo
Python
Docker
VSCode

Risultati del corso

Alla fine del corso i partecipanti avranno costruito:

- un sistema LLM locale funzionante
- una knowledge base interrogabile basata su documenti tecnici
- un modello AI specializzato su un dominio tecnico