

Titolo

Introduzione all'Intelligenza Artificiale generativa ed all'interazione con essa

Descrizione

La moderna Intelligenza Artificiale Generativa (IAG), o Generative Artificial Intelligence (AI) in inglese, è costituita da grandi modelli di linguaggio naturale (Large Language Models).

Questi sono il risultato dell'addestramento su enormi collezioni di dati (Big Data) di potenti e sofisticati algoritmi basati su moderne reti neurali (Transformers).

Si tratta di un ulteriore passo in avanti compiuto in un particolare tipo di apprendimento automatico (Machine Learning), noto come apprendimento profondo (Deep Learning), che governa da due decenni la moderna economia digitale del dato concepita ed attuata "in primis" dai grandi attori Big Tech (Google, Apple, Facebook/Meta, Amazon, Microsoft).

Il corso si pone l'obiettivo principale d'introdurre lo studente, in modo semplice, per passi e con l'ausilio di strumenti operativi visuali (KNIME), alla comprensione del funzionamento di base dell'IAG al fine di:

- imparare a coglierne in modo critico ed informato opportunità e limiti;
- disporre di strumenti teorico-concettuali propedeutici ad un impiego pratico proficuo dell'IAG in attività d'interesse centrate su produzione ed uso di linguaggio naturale.

Programma

Il corso è articolato secondo il seguente programma:

1. Introduzione al Machine Learning

- dati → algoritmi di apprendimento → modelli statistici previsionali
- apprendimento come minimizzazione di funzioni d'errore
- apprendimento supervisionato (modelli lineari, alberi decisionali)
- apprendimento non supervisionato (clustering)

2. Introduzione al Deep Learning

- le reti neurali per forme di apprendimento più complesse (modelli non lineari)
- propagazione avanti/indietro e "chain rule"
- potenza (e problemi) delle reti neurali

3. Reti Neurali Ricorrenti

- gestione di testo in linguaggio naturale con reti neurali specializzate
- da parole a numeri: il "word embedding"
- il problema dei gradienti nel processamento di sequenze linguistiche "lunghe"

4. Reti Neurali a Transformer

- reti neurali ricorrenti interconnesse per associare e generare testi
- il modello encoder/decoder (codifica/decodifica)
- il meccanismo di "attenzione" per una comprensione profonda di contesto e significato
- la generazione dei grandi modelli di linguaggio (es. ChatGPT, GPT4)

5. Interagire con grandi modelli di linguaggio proprietari ed open-source

- potenza e limiti dei grandi "risponditori statistici"
- elementi di prompt engineering

Previsti laboratori e dimostrazioni pratiche con KNIME Analytics Platform

Requisiti

È utile e vantaggiosa, ma non richiesta, una conoscenza (anche solo a livello di scuola secondaria) dei concetti principali di analisi matematica, algebra lineare, programmazione dei calcolatori elettronici.