

LEZIONE 3

Come funziona l'AI: LLM e machine learning

Corso di aggiornamento per docenti • I.I.S. "Tonino Guerra" – Cervia

Durata: 60 minuti • Animatore digitale



Struttura della lezione

In questa lezione esploreremo il funzionamento interno dell'intelligenza artificiale generativa: cosa accade davvero quando scriviamo un prompt, perché l'AI risponde in modo plausibile ma non sempre corretto, e cosa ci insegna tutto questo sulla sua utilità ed sui suoi limiti in classe.

Argomenti della lezione (60 min)

- Cos'è un LLM – Large Language Model
- AI generativa vs predittiva
- Addestramento su grandi dataset
- Token e previsione probabilistica
- Perché l'AI non capisce
- Perché sbaglia con sicurezza
- Differenze strutturali tra ChatGPT, Claude e Gemini
- Implicazioni didattiche del funzionamento tecnico

1. Cos'è un LLM (Large Language Model)

LLM è l'acronimo di Large Language Model, ovvero un modello linguistico di grandi dimensioni. È il motore che fa funzionare strumenti come ChatGPT, Claude, Gemini, Copilot e molti altri. Il termine "modello" indica un sistema matematico addestrato per svolgere un compito; "linguistico" perché il suo dominio principale è il linguaggio naturale; "large" perché è costruito su miliardi di parametri.

💡 Un LLM non è un database di risposte. È un sistema che impara schemi statistici dal linguaggio umano e li usa per generare testo nuovo, parola dopo parola.

Come nasce un LLM

- ▶ Viene esposto a enormi quantità di testo (libri, siti web, articoli, codice, forum...)
- ▶ Impara le relazioni statistiche tra le parole e i concetti
- ▶ Sviluppa rappresentazioni interne del linguaggio e della conoscenza
- ▶ Viene poi "affinato" con feedback umano per essere utile e sicuro

Un modello come GPT-4 ha circa 1.000 miliardi di parametri, connessioni matematiche tra concetti. Il cervello umano ha circa 100.000 miliardi di sinapsi, ma funziona in modo radicalmente diverso.

2. AI generativa vs AI predittiva

Non tutta l'AI funziona allo stesso modo. È utile distinguere due grandi famiglie, perché hanno finalità e logiche molto diverse.

	AI Predittiva	AI Generativa
Cosa fa	Classifica, prevede un output noto	Crea contenuti nuovi (testo, immagini, audio)
Esempi	Filtro antispam, rilevamento frodi, Netflix	ChatGPT, Claude, Midjourney, Suno
Output	Una categoria o un numero	Testo, immagine, audio, video
In classe	Analisi dei voti, suggerimenti adattivi	Scrittura, spiegazioni, materiali didattici

💡 Quando parliamo di AI a scuola ci riferiamo quasi sempre all'AI generativa. I docenti e gli studenti usano prevalentemente strumenti come ChatGPT o Claude — che sono LLM generativi.

3. Addestramento su grandi dataset

L'addestramento è la fase in cui il modello "impara" analizzando enormi quantità di testo. Non è un apprendimento consapevole: è un processo matematico di ottimizzazione.

Le fasi dell'addestramento

- ▶ **il modello analizza centinaia di miliardi di parole e impara a prevedere la parola successiva in una frase. Questo avviene su enormi cluster di GPU per settimane o mesi.** Pre-training.
- ▶ **il modello viene affinato su dati specifici (es. conversazioni utente-assistente) per adattarsi a compiti concreti.** Fine-tuning.
- ▶ **valutatori umani giudicano le risposte e il modello impara a produrre output più utili e sicuri.** RLHF (Reinforcement Learning from Human Feedback).

I dati di addestramento includono: libri digitalizzati, Wikipedia, siti web, Reddit, GitHub, articoli scientifici, contenuti social e molto altro. Questo spiega sia la ricchezza delle risposte sia i loro limiti e pregiudizi.

💡 Domanda per la classe: se un modello viene addestrato prevalentemente su testi in inglese scritti da uomini adulti di cultura occidentale, cosa ci aspettiamo riguardo ai suoi bias?

4. Token e previsione probabilistica

Capire il concetto di token è fondamentale per smettere di antropomorfizzare l'AI, cioè per smettere di immaginarla come un'entità che "pensa" o "capisce" davvero.

Cosa è un token

Un token è un'unità di testo, non esattamente una parola, ma un pezzo di parola (in media 3-4 caratteri in italiano). La frase "Come funziona l'AI" diventa circa 6 token.

Come funziona la generazione

- ▶ Il modello riceve i token di input (il tuo prompt)
- ▶ Calcola la probabilità di ogni possibile token successivo
- ▶ Sceglie il token più probabile (o ne campiona uno tra i più probabili)

- ▶ Aggiunge quel token all'output e ripete il processo
- ▶ Continua finché non produce un token di fine risposta

💡 L'AI non scrive la risposta "in testa" e poi la trascrive. La costruisce un pezzo alla volta, come un sistema di completamento automatico molto sofisticato. Questo spiega molti comportamenti strani.

Esempio pratico: se chiedete all'AI di continuare la frase "La capitale della Francia è..." risponderà "Parigi" perché statisticamente quella è la continuazione più probabile nel suo training. Non perché "sa" che Parigi è la capitale.

5. Perché l'AI non capisce

Questa è una delle distinzioni più importanti da trasmettere agli studenti. L'AI produce testo che sembra comprensione, ma non ha consapevolezza, intenzionalità, né modello del mondo.

Cosa sembra fare	Cosa fa davvero
Capire la domanda	Elaborare statisticamente i token di input
Ragionare	Generare token probabilisticamente coerenti
Ricordare	Usare il contesto della finestra di conversazione
Sapere se ha ragione	Non ha accesso a verifica esterna automatica
Avere opinioni	Produrre testo stilisticamente coerente con una "voce"

Il filosofo John Searle ha descritto questa distinzione con il famoso esperimento mentale della "stanza cinese": un sistema può manipolare simboli seguendo regole senza capire il significato di quei simboli. Gli LLM sono, in modo estremamente potente, sistemi di manipolazione simbolica.

6. Perché sbaglia con sicurezza

Una delle caratteristiche più pericolose degli LLM, soprattutto per gli studenti, è che sbagliano con un tono sicuro e autorevole. Questo fenomeno è noto come "allucinazione" (approfondito nella Lezione 4).

Le cause principali

- ▶ **non a essere accurato. La fluidità è la priorità interna.** Il modello è addestrato a produrre testo fluente e coerente,
- ▶ **anche quando dovrebbe dire "non lo so".** Non ha un meccanismo di "incertezza nativa": genera sempre qualcosa,
- ▶ **e il modello li ha assorbiti senza filtro critico.** I dati di addestramento contengono errori,
- ▶ **in conversazioni lunghe il modello può "dimenticare" informazioni precedenti.** La finestra di contesto è limitata:

💡 Per i docenti: non è una questione di "AI cattiva". È una proprietà strutturale del sistema. Per questo il pensiero critico e la verifica delle fonti sono competenze ancora più importanti nell'era dell'AI.

Domanda pratica da proporre in classe: "Come faccio a sapere se questa risposta dell'AI è affidabile?" Non esiste un segnale visivo di errore. L'AI non si "vergogna" di sbagliare.

7. Differenze strutturali tra ChatGPT, Claude e Gemini

Esistono molti modelli LLM, con architetture, dati di addestramento e filosofie di design differenti. Conoscere le differenze principali aiuta a scegliere lo strumento più adatto al compito.

Caratteristica	ChatGPT (OpenAI)	Claude (Anthropic)	Gemini (Google)
Modello attuale	GPT-5.2 (dic 2025)	Claude Sonnet 4.6 / Opus 4.6 (feb 2026)	Gemini 3.1 Pro (feb 2026)
Sviluppatore	OpenAI (+ Microsoft)	Anthropic	Google DeepMind
Punti di forza	Versatilità, ecosystem plugin, DALL·E, Memory tra sessioni	Scrittura, ragionamento complesso, codice, meno allucinazioni	Integrazione Google Workspace, multimodale, accesso al web in tempo reale
Finestra contesto	128K token (256K in modalità reasoning)	200K token (1M token in beta su Max)	1M token (2M token su piani avanzati)
Accesso gratis	Sì (GPT-5.2 Instant, con ads sul piano Go)	Sì (Claude Sonnet)	Sì — il piano gratuito più generoso tra i tre
Piani a pagamento	Go 8\$/mese, Plus 20\$/mese, Pro 200\$/mese	Pro 20\$/mese, Max 100–200\$/mese	AI Pro 20\$/mese (include Gemini 3.1 Pro)
Integrazione scuola	Microsoft Copilot / Teams / Office 365	claude.ai / API / Cowork (beta)	Google Workspace: Gmail, Docs, Drive, Classroom
Punti deboli	Allucinazioni frequenti, dati di addestramento non trasparenti	Ecosistema più piccolo, no generazione immagini native	Meno preciso nel ragionamento complesso rispetto a Claude
Filosofia design	Utilità generale massima	Safety-first, Constitutional AI	Multimodalità e integrazione nativa Google

💡 Non esiste il "miglior modello" in assoluto. La scelta dipende dal compito: per testi lunghi Claude, per ricerca con fonti Perplexity, per integrazione Microsoft Copilot, per Google Workspace Gemini.

8. Implicazioni didattiche del funzionamento tecnico

Comprendere come funziona l'AI non è solo curiosità tecnica: ha conseguenze dirette su come la utilizziamo e come insegnamo ai nostri studenti a usarla responsabilmente.

Cosa cambia nella didattica
• Insegnare la verifica: l'AI sbaglia con sicurezza → serve sempre il fact-checking • Prompt come competenza: la qualità dell'input determina la qualità dell'output • No all'antropomorfizzazione: l'AI non "capisce", non "sa", non "vuole" • Contestualizzare gli errori: un errore AI non è un'eccezione, è strutturale • Usare l'AI come specchio critico: confrontare le risposte, non accettarle passivamente • Riflettere sui dati: chi ha scritto i testi su cui l'AI si è addestrata? Quali voci mancano?

Attività pratica suggerita (20 min in classe)

- ▶ Chiedere alla stessa AI la stessa domanda due volte → le risposte differiscono (probabilità, non verità)
- ▶ Inventare un personaggio storico inesistente e chiedergli informazioni → osservare l'allucinazione
- ▶ Fare una domanda con risposta nota errata (es. "Quanti giorni ha febbraio in ogni anno?") → osservare se l'AI corregge o asseconda

💡 Il valore di questa lezione non è tecnico ma epistemologico: aiuta studenti e docenti a sviluppare un rapporto consapevole con uno strumento che produce testo autorevole ma non infallibile.

Riepilogo della lezione

Concetto	Messaggio chiave
LLM	Sistema matematico che impara schemi dal linguaggio, non un database
Generativa vs predittiva	L'AI che usiamo a scuola crea testo nuovo, non recupera dati preesistenti
Addestramento	Apprende da enormi corpus di testo; assorbe conoscenza e pregiudizi
Token	La risposta viene costruita un pezzo alla volta, statisticamente
Non capisce	Manipola simboli con grande efficacia senza vera comprensione
Sbaglia con sicurezza	La fluidità non implica l'accuratezza: il fact-checking è sempre necessario
Modelli diversi	ChatGPT, Claude, Gemini hanno punti di forza diversi: scegliere con criterio
Didattica	Conoscere il funzionamento rende l'uso critico, consapevole e formativo

? Spunti per la discussione finale

- ▶ Dopo questa lezione, come descrivereste l'AI a un vostro studente di 13 anni?
- ▶ Qual è il rischio principale di usare l'AI senza conoscerne il funzionamento?
- ▶ Come cambia il valore del "pensiero critico" degli studenti nell'era degli LLM?
- ▶ Quale di questi concetti vi sembra più urgente trasmettere in classe?