

MASSIMILIANO TARQUINI

MATTONE DOPO MATTONE

Comprendere gli LLM

Una guida non tecnica
ai modelli linguistici
di grandi dimensioni



MATTONE DOPO MATTONE

Comprendere i modelli linguistici di grandi dimensioni

Una guida non tecnica

Massimiliano Tarquini

AiOne

Edizione italiana revisionata — 5 ottobre 2026

Sommario

Parte I — Fondamenti	5
Introduzione	5
Che cosa sono i modelli linguistici di grandi dimensioni	5
Come funzionano senza entrare troppo nei dettagli	6
Parte II — Capacità e interazione	9
Che cosa possono fare gli LLM	9
Parlare con gli LLM — Perché i prompt contano e come vengono progettate le interfacce	10
Capacità emergenti — Che cosa imparano i modelli da soli	12
Oltre le domande — Come usare il prompting per sfruttare le capacità emergenti	16
Dalle parole alle capacità — Esempi pratici di prompting emergente	17
Parte III — Etica, limiti e società	20
Domande importanti — Che cosa dovremmo chiederci sugli LLM	20
Il futuro degli LLM — Rischi e limiti di ciò che non possono e non dovrebbero fare	21
Usare gli LLM con criterio	21
IA centrata sulle persone e spiegabilità — Fidarsi di ciò che si può comprendere	24
Chiedere ai modelli di spiegarsi	25
Esempi pratici di spiegabilità in azione	26
Miti e realtà — Che cosa gli LLM non possono fare anche quando sembra che ne siano capaci	28
Quando la fiducia eccessiva incontra l'intelligenza artificiale	31
Una riflessione finale	35
Parte IV — Applicazioni ed esempi	36
Esempi dal mondo reale	36
Strumenti quotidiani alimentati discretamente dall'IA	37
Opportunità e vantaggi — Perché gli LLM possono essere davvero utili	37
Storie inattese dal mondo dei modelli linguistici	38

Sommario

Parte V — Dall'intelligenza generale a quella specializzata	40
Dal generale allo specialistico — Come gli LLM imparano nuove competenze	40
Gli LLM come macchine della memoria — Recuperare conoscenze ed esperienza umana	41
Un intero mondo di modelli linguistici — Dai giganti ai piccoli aiutanti	42
Non soltanto parole — Quando i modelli linguistici vedono e comprendono le immagini	44
Parte VI — Infrastruttura e sostenibilità	46
Potenza e prezzo — Il costo nascosto dell'intelligenza	46
Confrontare i modelli di IA — Giganti del cloud ed esperti locali	47
Parte VII — Appendici	49
Prima di chiedere — Una checklist per usare gli LLM con criterio	49
Conclusione	49
Glossario	50
Vuoi saperne di più — Risorse da esplorare	51

Parte I — Fondamenti

Nota sull'edizione italiana revisionata. Questo volume conserva il percorso e gli argomenti della guida originale, integrandoli con correzioni tecniche, riferimenti controllabili e attività guidate. I racconti non documentati sono esplicitamente presentati come scenari illustrativi. La traduzione integrale precedente rimane disponibile separatamente. Per seguire gli esempi basta saper leggere una breve tabella e confrontare due testi; non servono conoscenze di programmazione. L'obiettivo è imparare a scegliere un impiego del modello, verificarne il risultato e riconoscere quando occorre un'altra soluzione.

Introduzione

L'intelligenza artificiale non appartiene più soltanto alla fantascienza: fa parte della nostra vita quotidiana. Potresti averci già interagito senza rendertene conto, quando il telefono suggerisce parole mentre scrivi, quando un chatbot risponde alle tue domande o quando uno strumento ti aiuta a riscrivere un'email.

Al centro di molti di questi strumenti intelligenti ci sono i Large Language Models, in breve LLM: i modelli linguistici di grandi dimensioni. Questi sistemi di intelligenza artificiale stanno cambiando il modo in cui lavoriamo, impariamo e comunichiamo. Ma che cosa sono esattamente? Come funzionano? E perché suscitano tanta attenzione?

Questa guida ti aiuta a capire che cosa sono gli LLM senza ricorrere al gergo tecnico. Che tu sia curioso, preoccupato o desideroso di restare al passo con le ultime tendenze, troverai risposte semplici e chiare.

Il documento ti accompagnerà alla scoperta di questi argomenti:

- Che cosa sono gli LLM, come vengono addestrati e che cosa possono fare.
- Come interagiamo con loro attraverso prompt in linguaggio naturale.
- Perché a volte ci sorprendono con capacità inattese.
- Quali rischi e limiti presentano: bias, allucinazioni e mancanza di comprensione.
- Come usarli con criterio, con una checklist pratica.
- Dove vengono impiegati oggi, dalla sanità all'istruzione e alla creatività.
- Le nuove frontiere, come i modelli multimodali che vedono e comprendono le immagini.
- Quanto costano in termini di energia, denaro e potenza di calcolo.

Non occorre essere sviluppatori per capire tutto questo. Bastano un po' di curiosità e il desiderio di confrontarsi con una delle tecnologie più potenti del nostro tempo, in modo responsabile, creativo e con spirito critico.

Che cosa sono i modelli linguistici di grandi dimensioni

Immagina di voler insegnare a un computer a comprendere e usare il linguaggio. Non soltanto con un dizionario o un libro di grammatica, ma lasciandogli leggere

quasi tutto: romanzi, articoli di giornale, pagine di Wikipedia, post sui social, manuali tecnici, poesie e perfino sceneggiature cinematografiche.

È così che vengono addestrati i modelli linguistici di grandi dimensioni, gli LLM. Vengono esposti a quantità enormi di testo, così da imparare come le persone parlano e scrivono davvero nelle situazioni più diverse.

Ma ecco l'aspetto sorprendente:

Molti LLM generativi apprendono inizialmente a prevedere il token successivo. Ulteriori fasi di addestramento possono adattarli a istruzioni e compiti.

Un token è un'unità di testo, spesso una parola o un suo frammento. La previsione utilizza rappresentazioni apprese delle relazioni fra elementi: il meccanismo è semplice da descrivere, ma le capacità che ne risultano richiedono prove. Non deduciamo dal solo obiettivo di addestramento né una comprensione umana né l'assenza di capacità di ragionamento.

Per esempio:

- Se scrivi «Il cielo è pieno di...», potrebbe suggerire stelle, nuvole o aerei.
- Se chiedi «Puoi spiegarmi la fotosintesi?», attinge a tutto ciò che ha letto per costruire una risposta che appare utile e coerente.
- Chiedigli di scrivere una poesia, riassumere un articolo, tradurre una frase o persino aiutarti con il codice: spesso è in grado di farlo.

Perché funziona? Perché il linguaggio segue schemi ricorrenti. Più esempi vede il modello, più diventa bravo a indovinare ciò che viene dopo; queste previsioni iniziano ad assomigliare molto al ragionamento, alla memoria o alla creatività.

Ecco perché gli LLM vengono spesso descritti come motori statistici del linguaggio. Non conoscono i fatti come li conoscono gli esseri umani, ma sono straordinariamente abili nell'imitare la comunicazione umana sulla base di ciò che hanno letto.

Ecco alcuni esempi di ciò che possono fare:

- Scrivere lettere, post per blog, favole della buonanotte e discorsi.
- Conversare: rispondere alle domande, sostenere conversazioni e persino interpretare un ruolo.
- Tradurre testi fra lingue diverse, compresi usi informali e gergali.
- Riassumere documenti lunghi, verbali di riunioni e articoli accademici.
- Aiutare con il codice: spiegare concetti di programmazione o generare frammenti di codice.

Vengono utilizzati negli assistenti virtuali, nei bot per il servizio clienti, negli strumenti di scrittura, nelle applicazioni didattiche e nei progetti creativi. Quindi, anche se un LLM è soltanto un predittore di parole molto avanzato, l'esposizione al linguaggio umano gli permette di svolgere una varietà sorprendente di compiti, spesso in modi che sembrano del tutto naturali.

Come funzionano senza entrare troppo nei dettagli

Supponiamo che tu stia scrivendo a un amico e digiti:

«Mi sento davvero...»

Prima ancora che tu finisca la frase, il telefono potrebbe suggerire parole come stanco, felice o emozionato. È un piccolo esempio di ciò che fa un modello linguistico: cerca di indovinare la parola successiva sulla base di quelle che hai già scritto.

Ora immagina di farlo non soltanto con poche frasi, ma con miliardi di frasi, in stili diversi e su argomenti differenti. È ciò che fanno gli LLM. Diventano estremamente bravi a prevedere ciò che viene dopo nel linguaggio.

Al centro di questo processo c'è un tipo di intelligenza artificiale chiamato deep learning, o apprendimento profondo, che utilizza reti neurali. Sono sistemi digitali vagamente ispirati al funzionamento del cervello umano, composti da strati che elaborano le informazioni passo dopo passo. Ogni strato cerca schemi ricorrenti:

- Uno potrebbe occuparsi dell'ortografia o della struttura delle parole.
- Un altro potrebbe concentrarsi sulla grammatica.
- Un altro ancora potrebbe cercare elementi di contesto più profondi, come il tono o l'intenzione.

La divisione fra ortografia, grammatica e contesto è un'analogia, non una mappa letterale degli strati: le rappresentazioni sono distribuite e non assegniamo a ogni strato un solo compito. I parametri vengono regolati durante l'addestramento. Con il tempo, il modello impara a formulare previsioni sempre più accurate e sempre più simili a quelle umane.

Un'analogia semplice — I mattoncini LEGO

Immagina di costruire una frase con mattoncini LEGO. Il modello impara quali mattoncini tendono a stare bene insieme: «burro di arachidi» compare spesso con «marmellata», «medico» potrebbe comparire con «ospedale» e «clima» con «cambiamento».

Non conosce il significato di queste parole come lo conoscono gli esseri umani, ma sa come tendono a comparire insieme.

Perché le dimensioni contano

Aumentare dati e capacità del modello può migliorare le prestazioni, ma il risultato dipende anche dalla qualità dei dati, dall'architettura, dall'addestramento e dal compito. Un modello più grande non è automaticamente migliore per ogni uso. Ecco perché li chiamiamo modelli linguistici di grandi dimensioni: sono composti da miliardi di parametri, simili a piccole manopole che il modello regola per migliorare le proprie previsioni.

Il risultato? Un sistema che può:

- Scrivere un comunicato stampa in pochi secondi.
- Rispondere a domande complesse.
- Simulare una conversazione.
- Perfino spiegare battute o giochi di parole, almeno qualche volta.

Ma ricorda:

Concetto chiave — Prevedere non significa soltanto completare frasi. Un modello generativo produce token, cioè unità di testo che possono essere parole o loro parti, usando rappresentazioni apprese. Questo meccanismo può sostenere compiti complessi; da solo non dimostra coscienza, esperienza personale o affidabilità fattuale.

Prova guidata — Completare non equivale a verificare. Devi conoscere un orario, ma il testo disponibile dice soltanto «Il negozio apre alle». Un completamento plausibile non risolve il problema. Fornisci invece «Il negozio apre alle 9» e chiedi di estrarre l'orario senza aggiungere dati: il risultato atteso è 9. Verifica sul testo, poi togli l'orario e controlla che venga segnalata l'assenza. Il confronto separa generazione e riscontro; riusalo per estrarre dati espliciti, senza dedurne fatti esterni.

Parte II — Capacità e interazione

Che cosa possono fare gli LLM

I modelli linguistici di grandi dimensioni sono come i coltellini svizzeri del mondo digitale: versatili, potenti e ricchi di strumenti inattesi. L'addestramento alla previsione dei token e le fasi successive possono sostenere molte attività, da valutare sul caso concreto.

Ecco alcune delle cose che possono fare, con esempi concreti.

Scrittura e revisione

Hai bisogno di aiuto per preparare un'email? Il modello può suggerire un tono cortese e professionale. Il blocco dello scrittore ti mette in difficoltà? Può proporti alcune frasi iniziali. Vuoi riscrivere un documento in un linguaggio più semplice? Bastano pochi secondi.

Esempio: «Gentile Professore, Le scrivo per richiedere una proroga...» → l'LLM completa il testo con un linguaggio formale e rispettoso.

Riassumere testi lunghi

Non hai tempo per leggere una relazione di dieci pagine? Chiedi al modello un breve riassunto. Stai rivedendo appunti di riunioni o documenti sulle politiche di un'organizzazione? Il modello può estrarne i punti principali.

Esempio: «Riassumi questo articolo sul cambiamento climatico in tre punti».

Traduzione e aiuto con il linguaggio

Tradurre testi fra lingue diverse con una fluidità sorprendente. Riformulare una frase per renderla più formale, informale o chiara. Spiegare modi di dire, espressioni e perfino il gergo.

Esempio: «Che cosa significa "rompere il ghiaccio"?» → il modello lo spiega come un insegnante cordiale.

Conversazione e assistenza

Simulare un assistente utile per il servizio clienti o il supporto tecnico. Rispondere alle domande più comuni su un prodotto, una politica o una procedura.

Esempio: un chatbot basato su un LLM può guidare gli utenti nella configurazione del Wi-Fi o nella prenotazione di un appuntamento.

Tutoraggio e spiegazioni

Aiutare gli studenti a comprendere argomenti complessi riformulandoli in termini più semplici. Generare domande per quiz o esempi su cui esercitarsi. Fare da «compagno di studio» che spiega le cose passo dopo passo.

Esempio: «Spiegami la gravità come se avessi dieci anni» → il modello utilizza un'analogia accessibile e amichevole.

Programmazione e aiuto tecnico

Suggerire soluzioni a problemi di programmazione. Generare piccoli frammenti di codice. Aiutare a individuare gli errori o a spiegarli.

Esempio: «Scrivi una funzione Python che controlli se un numero è primo».

Comprendere i dati con la configurazione giusta

Quando sono collegati a fonti di dati o a dashboard, gli LLM possono descrivere le tendenze in linguaggio naturale. Possono aiutare a trasformare i numeri in un racconto.

Esempio: «Spiegami in parole semplici che cosa significa questo grafico».

Come puoi vedere, gli LLM non sono riservati agli esperti di tecnologia: stanno già aiutando scrittori, studenti, professionisti, operatori sanitari, giornalisti e artisti. Non sostituiscono le competenze umane, ma possono amplificarle, agendo come copiloti digitali.

Parlare con gli LLM — Perché i prompt contano e come vengono progettate le interfacce

I modelli linguistici di grandi dimensioni non utilizzano pulsanti, menu o comandi predefiniti. Rispondono invece a input in linguaggio naturale, chiamati anche prompt. Non è soltanto una funzione che rende l'uso più semplice: è un principio fondamentale di progettazione.

A differenza del software tradizionale, che segue istruzioni rigide, gli LLM operano in uno spazio flessibile e probabilistico. Un prompt agisce sia come input sia come contesto: definisce il compito, stabilisce il tono e spesso suggerisce il formato atteso della risposta.

Perché utilizzare interfacce basate sui prompt

I prompt permettono un'interazione aperta. È essenziale, perché gli LLM sono sistemi di uso generale. Aniché costruire una nuova applicazione per ogni funzione, possiamo riutilizzare lo stesso modello per compiti diversi, semplicemente cambiando il prompt.

Per esempio:

«Scrivi un'email formale per richiedere un incontro».

«Spiega le leggi della termodinamica in termini semplici».

«Riassumi questo documento sulle politiche aziendali in cinque punti».

In ogni caso, l'utente definisce in linguaggio naturale che cosa deve fare il modello.

Il prompting come livello di controllo

In termini tecnici, il prompting è un livello di interfaccia flessibile. Non modifica il modello sottostante, ma ne orienta il comportamento. Un prompt ben costruito può:

- Assegnare un ruolo, per esempio «Sei un consulente legale...».
- Stabilire un compito, per esempio «Prepara una clausola contrattuale sulla riservatezza».
- Definire lo stile o il formato, per esempio «Scrivi un elenco puntato di non più di cento parole».

Ecco perché il prompting viene spesso descritto come una forma di programmazione con le parole: attraverso un linguaggio che guida gli output del modello, anziché attraverso il codice.

Dall'uso occasionale all'ingegneria dei prompt

I prompt semplici funzionano per i compiti quotidiani. Ma quando gli utenti iniziano a spingersi oltre, combinando più vincoli, utilizzando messaggi di sistema o concatenando istruzioni, il prompting diventa una competenza a sé.

Da qui nasce il prompt engineering, o ingegneria dei prompt: la pratica di costruire input che ottimizzino la pertinenza, la precisione o la creatività dell'output.

Un buon prompting non si basa sui tentativi casuali. Richiede di capire come gli LLM interpretano e ponderano gli indizi testuali, gestire l'ambiguità per evitare allucinazioni e strutturare gli input per ragionamenti in più passaggi o compiti specifici di un dominio.

Interfacce costruite intorno al linguaggio

Che si tratti di chatbot, assistenti integrati, ambienti di sviluppo o dashboard, la tendenza è chiara: il linguaggio stesso sta diventando l'interfaccia. Una progettazione basata sui prompt riduce gli ostacoli, aumenta l'accessibilità e si adatta al modo in cui gli esseri umani comunicano naturalmente.

Questo non significa che gli elementi delle interfacce grafiche scompariranno. Molti strumenti combinano già input basati su moduli e campi per i prompt, offrendo agli utenti sia struttura sia libertà.

Prova guidata — Trasformare una richiesta vaga. Vuoi una mail di conferma, ma «Scrivi una mail» lascia scopo e destinatario indefiniti. Prova «Scrivi una mail a un collega per confermare l'incontro di martedì alle 10; tono cordiale; massimo 80 parole; non inventare luogo o nomi». Controlla tutti i vincoli. Come variante, cambia destinatario e mantieni il contenuto: il tono dovrebbe adattarsi senza alterare data e ora. Riusa il metodo per compiti testuali con vincoli; non aspettarti che un prompt dettagliato garantisca fatti veri.

Capacità emergenti — Che cosa imparano i modelli da soli

È qui che le cose diventano davvero affascinanti. Quando i modelli linguistici di grandi dimensioni crescono e diventano più complessi, iniziano a fare cose che nessuno ha insegnato loro esplicitamente.

Molti modelli generativi vengono inizialmente addestrati a prevedere il token successivo; possono seguire altre fasi di addestramento, con esempi di istruzioni, preferenze e compiti. Durante questo percorso si sviluppano capacità che vengono misurate mediante prove specifiche. È come se il modello iniziasse a «cogliere» più aspetti del mondo di quanto ci si aspettasse.

Che cosa sono le capacità emergenti

Si parla di capacità emergenti quando una valutazione rileva un comportamento nei modelli più grandi che non era rilevabile nei più piccoli. Non è una legge per cui ogni abilità compare oltre una soglia universale: contano anche dati, addestramento, prompt e metrica, cioè la regola usata per assegnare un punteggio.

Ecco alcuni esempi:

- Risolvere indovinelli o rompicapi logici: proponi al modello un enigma e potrebbe trovare la risposta.
- Scrivere codice funzionante: descrivi un compito in linguaggio naturale e il modello genera codice che funziona.
- Risolvere problemi matematici espressi a parole: può seguire la logica di una domanda e trovare la formula giusta per risolverla.
- Riconoscere sarcasmo o significati nascosti: coglie indizi sottili e il tono del linguaggio.
- Interpretare ruoli o immaginare prospettive: può simulare punti di vista differenti o «fingere» di essere un insegnante, un medico o un personaggio storico.

Perché succede

Quando addestri un modello su quantità enormi di testo umano, milioni di conversazioni, storie, spiegazioni e dibattiti, il modello inizia a generalizzare. Significa che può applicare ciò che ha visto in un contesto a una situazione nuova e sconosciuta.

Possiamo descrivere questo comportamento come «intuizione artificiale» soltanto in senso metaforico: il modello formula ipotesi a partire dalle rappresentazioni apprese. Per valutarle occorre controllare se risolvono il problema, senza inferire un'esperienza interiore.

È simile a ciò che accade a una persona che legge molto e riesce spesso a rispondere a domande o a spiegare cose che nessuno le ha insegnato formalmente, perché ha assimilato tanti esempi.

Una sorta di intuizione artificiale

La «teoria della mente» indica la capacità di attribuire ad altri credenze, conoscenze e intenzioni. Alcune prove linguistiche chiedono di prevedere la risposta di un personaggio che possiede informazioni diverse dalle nostre. Un successo in queste prove va distinto dalla dimostrazione che il modello possieda la stessa capacità umana in ogni situazione.

Come abbiamo scoperto le capacità emergenti

Molte delle capacità sorprendenti degli LLM non sono state progettate esplicitamente. Nessuno ha scritto codice per insegnare loro a risolvere indovinelli, scrivere programmi funzionanti o tradurre lingue poco diffuse. Eppure, durante i test e l'esplorazione, modelli come GPT-3 e GPT-4 hanno iniziato a mostrare comportamenti che andavano ben oltre la semplice previsione delle parole.

Come ce ne siamo accorti? In parte per caso, in parte per curiosità. Ricercatori e utenti hanno iniziato a sperimentare prompt diversi, chiedendo al modello di «spiegare come un insegnante», di «ragionare passo dopo passo» o di risolvere rompicapi logici. Le risposte potevano essere coerenti e accurate su alcune prove. La generalizzazione delle rappresentazioni apprese contribuisce a queste prestazioni; la fluidità della risposta, da sola, non permette di ricostruire il processo né di stabilirne l'affidabilità su un'altra prova.

Wei e colleghi hanno discusso queste capacità nel 2022. Schaeffer, Miranda e Koyejo hanno poi mostrato che alcuni salti apparenti dipendono dalla metrica: un miglioramento graduale può sembrare improvviso se il punteggio distingue soltanto successo e fallimento. Il dibattito invita a controllare come misuriamo le prestazioni, oltre alle dimensioni del modello.

Riferimento essenziale — Emergenza e misurazione. Wei e colleghi, *Emergent Abilities of Large Language Models* (2022), descrivono risultati nei quali la scala è associata a nuove prestazioni. Schaeffer, Miranda e Koyejo, *Are Emergent Abilities of Large Language Models a Mirage?* (2023), discutono l'effetto delle metriche. I due lavori aiutano a separare l'abilità misurata dalla spiegazione del suo apparire; non stabiliscono una soglia valida per tutti i modelli. Riferimenti completi nelle risorse finali.

Uno scorcio di teoria della mente

In alcuni studi, i modelli di grandi dimensioni hanno dimostrato la capacità di prevedere ciò che un'altra persona potrebbe pensare: una facoltà chiamata teoria della mente. Per esempio, se un personaggio di una storia non sa qualcosa che il lettore conosce, il modello potrebbe comunque mantenere limitata la conoscenza di quel personaggio.

Questo significa che il modello è cosciente? No.

Mostra però quanto lontano possa arrivare il riconoscimento statistico degli schemi quando viene esteso a dimensioni enormi.

Le capacità emergenti sono una parte di ciò che rende gli LLM al tempo stesso entusiasmanti e misteriosi. Sollevano grandi domande sull'intelligenza, sulla creatività e su ciò che le macchine potrebbero essere in grado di fare in futuro.

Queste funzioni non sono codificate esplicitamente: emergono dall'apprendimento di schemi in volumi enormi di testo umano.

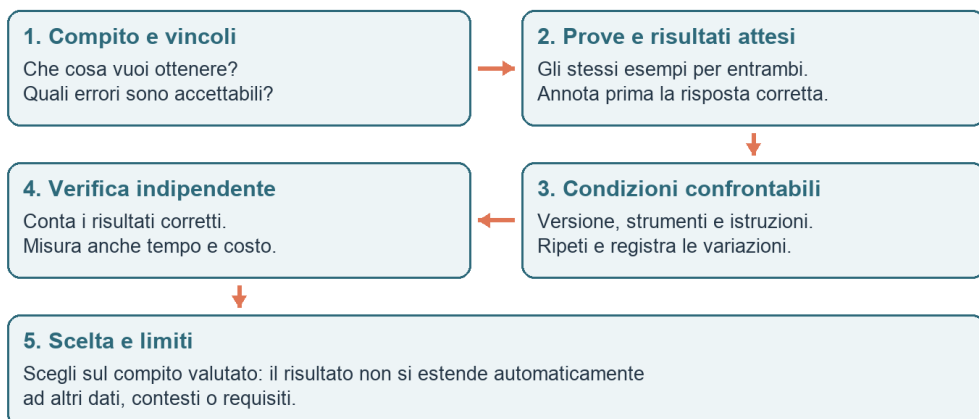
Capacità	Che cosa può fare il modello	
Risolvere problemi logici e matematici	Utilizza un ragionamento passo dopo passo per risolvere enigmi, equazioni o problemi espressi a parole, con accuratezza crescente.	Mostra forme iniziali di ragionamento simbolico e la capacità di generalizzare schemi logici.
Scrivere codice a partire dal linguaggio naturale	Genera codice funzionante in Python, JavaScript e altri linguaggi, spiega gli errori e migliora gli script esistenti.	Trasforma il linguaggio in logica eseguibile, collegando il ragionamento umano a quello delle macchine.
Comprendere contenuti multimodali	Interpreta immagini, grafici o schermate e li spiega a parole, come avviene per esempio con GPT-4V e Gemini.	Collega il testo alla visione, un aspetto cruciale per istruzione, medicina e ricerca.
Ragionare attraverso più scambi	Mantiene coerenza logica e contesto durante una conversazione o un'analisi in più passaggi.	Rende possibili dialoghi prolungati, pianificazione e scomposizione dei compiti.
Riconoscere tono, emozioni e sarcasmo	Riconosce umorismo, ironia e tono emotivo nella conversazione.	Rende l'interazione fra persone e IA più naturale e sensibile al contesto.
Tradurre lingue poco diffuse o mescolate	Gestisce lingue con poche risorse disponibili, dialetti e code-switching, cioè l'alternanza fra due lingue.	Amplia l'accesso e l'inclusione nelle diverse comunità linguistiche.
Generare dati strutturati	Produce tabelle, JSON o output basati su schemi direttamente dal testo naturale.	È essenziale per automazione, pipeline di dati e integrazione con strumenti aziendali.
Inferire il contesto nascosto nei test di teoria della mente	Prevede ciò che un personaggio di una storia sa o crede, distinguendolo da ciò che sa o crede un altro.	È un primo segnale di modellazione del contesto, della prospettiva e dell'intenzione.
Simulare ruoli o esperti	Assume personaggi come insegnante, medico o analista e risponde di conseguenza.	Rende l'IA adattabile agli obiettivi e agli stili di comunicazione degli utenti.
Spiegare una risposta attraverso passaggi intermedi	Può proporre una sequenza di passaggi da controllare.	Migliora la leggibilità della soluzione, senza certificare il processo interno che l'ha prodotta.
Generare ipotesi e analogie	Suggerisce spiegazioni possibili o confronti creativi.	Sostiene l'ideazione, la ricerca e l'istruzione.

Capacità	Che cosa può fare il modello	
Autocorreggersi e riflettere	Rivede o critica il proprio output precedente quando gli viene richiesto.	Introduce un perfezionamento iterativo e un ragionamento sul ragionamento.
Seguire prompt complessi in più passaggi	Esegue sequenze come «analizza → ragiona → riassumi → consiglia».	Costituisce la base dell'uso di strumenti e del comportamento agentic.

Cose sorprendenti che gli LLM hanno imparato a fare

Le prestazioni dei modelli cambiano nel tempo, ma un confronto richiede compiti e condizioni definiti. Il radar della guida originale attribuiva punteggi a GPT-4 e GPT-5 senza dichiarare prove, scala o metodo. In questa edizione lo sostituiamo con lo schema seguente: serve a costruire un confronto verificabile, senza introdurre nuove misure.

Un confronto che puoi verificare



Schema di metodo • Nessun punteggio attribuito ai modelli

Figura 1 — Come confrontare due modelli sullo stesso compito. Lo schema descrive un metodo e non attribuisce punteggi ai prodotti.

La figura collega cinque passaggi: scegliere il compito, preparare esempi con un risultato atteso, mantenere le condizioni confrontabili, controllare gli output e decidere in base a qualità, costo e tempo. Un benchmark è un insieme di prove con una regola di valutazione; misura la prestazione su quelle prove, non un'intelligenza generale su una scala unica.

Esempio svolto — Un confronto utile. Devi estrarre data e importo da dieci ricevute fittizie. Prima annoti a mano le risposte corrette; poi invii gli stessi testi ai due modelli, con gli stessi strumenti e lo stesso formato richiesto. Conti i campi corretti, registri il tempo e ripeti la prova, perché gli output possono variare. Se un modello è più accurato ma troppo lento per il tuo servizio, la scelta resta aperta. Come variante, aggiungi una ricevuta senza data: un modello affidabile per questo compito deve segnalare il dato mancante, invece di inventarlo. Riusa il confronto soltanto per input e condizioni simili; una prova di estrazione non dimostra capacità di consulenza medica.

Oltre le domande — Come usare il prompting per sfruttare le capacità emergenti

Il prompting non consiste nel dare ordini, ma nel sintonizzare una risonanza. Quando scriviamo a un modello linguistico, non gli chiediamo soltanto di produrre parole: modelliamo il modo in cui collega i molti livelli di conoscenza e ragionamento che già possiede. Un prompt può favorire ragionamento, analogia, creatività e autocorrezione sul compito; non garantisce che una capacità sia presente né che venga esercitata correttamente.

Un prompt può fornire istruzioni, fatti ed esempi nuovi nel contesto della risposta. Questo adattamento non implica che i parametri vengano aggiornati: distinguiamo ciò che il sistema usa nella conversazione da ciò che ha appreso durante l'addestramento.

Come i prompt possono attivare capacità emergenti

I comportamenti emergenti richiedono spesso indizi contestuali espliciti. Lo stesso modello può rispondere in modo superficiale o brillante, a seconda di come viene formulato il compito. Costruendo prompt che rispecchiano il modo in cui gli esseri umani affrontano il ragionamento, possiamo aiutare il modello a far emergere le sue capacità più complesse.

Attivazione della catena di pensiero

Chiedere al modello di «ragionare passo dopo passo», «spiegare il proprio ragionamento» o «giustificare la risposta» lo incoraggia a seguire passaggi intermedi, favorendo una soluzione articolata che possiamo controllare, senza assumere che il testo riveli i meccanismi interni del modello.

Ragionamento basato sul ruolo

Quando diciamo «Sei un analista delle politiche pubbliche che valuta una nuova normativa» oppure «Agisci come un ricercatore medico che interpreta dati», guidiamo il modello verso un lessico, un tono e un'impostazione analitica specialistici, concentrando di fatto le sue competenze interne.

Pensiero comparativo e analogico

Prompt come «Confronta questi due casi» o «Trova un'analogia che spieghi questo concetto» attivano la capacità del modello di mettere in corrispondenza schemi diversi, uno dei suoi punti di forza emergenti più profondi.

Autoriflessione e revisione

Istruzioni come «Rivedi la tua risposta precedente e migliorane la chiarezza» o «Individua possibili errori nel tuo ragionamento» sollecitano una revisione che può migliorare la risposta o introdurre nuovi errori. Confronta ogni modifica con il testo e con fonti indipendenti.

Progettare i prompt come percorsi cognitivi

Come metafora, un prompt può essere immaginato come un percorso nello spazio delle rappresentazioni apprese dal modello. Lo spazio latente è una rappresentazione numerica delle relazioni fra elementi; il prompt non è un comando che ci permette di ispezionarla direttamente. Un prompt minimo o vago lascia il modello a vagare; un prompt preciso e articolato agisce come un sentiero ben segnalato che collega ragionamento, memoria ed espressione.

Un buon prompting sfrutta questa possibilità incorporando struttura, prospettiva e intenzione nel linguaggio naturale:

«Ragioniamo con attenzione, passo dopo passo, considerando sia le implicazioni etiche sia quelle pratiche prima di arrivare a una conclusione».

Una formulazione di questo tipo invita il modello a coordinare più capacità emergenti contemporaneamente: ragionamento logico, inquadramento morale e sintesi, producendo risposte che appaiono sia analitiche sia sensibili al contesto umano.

Ragionamento collaborativo — La persona nel ciclo

Sfruttare le capacità emergenti attraverso il prompting trasforma infine l'interazione fra persona e IA in una forma di ragionamento collaborativo. Il modello amplifica l'intuizione linguistica e concettuale; la persona fornisce direzione, vincoli e valutazione. Ogni scambio affina il processo emergente, mantenendolo coerente e orientato a uno scopo.

L'idea è semplice ma essenziale: gli LLM possono ragionare, generalizzare e correggere alcuni errori; una guida esplicita può aiutare, ma non è sempre necessaria né sufficiente. In questo senso, il prompting non è un trucco: è una forma di comunicazione che insegna al modello a pensare con noi, non soltanto per noi.

Dalle parole alle capacità — Esempi pratici di prompting emergente

Ogni prompt ben progettato è più di una domanda: è un segnale che attiva determinati comportamenti cognitivi nel modello. Gli esempi seguenti mostrano

come diversi tipi di prompt possano stimolare deliberatamente specifiche capacità emergenti, facendo affiorare il ragionamento, la creatività o l'adattabilità più profondi del modello.

Esempio di prompt	Capacità emergenti stimulate	Effetto sul comportamento del modello
«Ragioniamo passo dopo passo prima di dare la risposta finale».	Sequenziamento logico e ragionamento causale attraverso la catena di pensiero.	Incoraggia un ragionamento articolato e spiegazioni coerenti, passo dopo passo, anziché risposte immediate.
«Sei un analista di sicurezza informatica che esamina un log di rete sospetto. Individua le possibili minacce e giustifica le tue conclusioni».	Ragionamento basato sul ruolo, adattamento al dominio e concentrazione analitica.	Il modello adotta un linguaggio specialistico, applica un'analisi strutturata e filtra i dettagli irrilevanti.
«Confronta il modo in cui le teorie di Darwin e Lamarck spiegano l'evoluzione, evidenziando un punto di forza e un limite per ciascuna».	Ragionamento comparativo, astrazione e sintesi equilibrata.	Attiva un'analisi per contrasto e un inquadramento concettuale di prospettive diverse.
«Riscrivi la tua risposta precedente in termini più semplici, adatti a un lettore di dieci anni».	Autoriflessione, metacognizione e ricodifica contestuale.	Il modello rivaluta il proprio output, misurandone chiarezza e accessibilità.
«Trova un'analogia che spieghi le reti neurali a qualcuno che non ha mai studiato l'IA».	Ragionamento analogico, corrispondenze metaforiche e trasferimento creativo.	Coinvolge il riconoscimento di schemi fra domini differenti, per produrre spiegazioni intuitive.
«Dato questo testo, estrai l'argomento centrale ed elenca due controargomentazioni che potrebbero metterlo in discussione».	Ragionamento critico, pensiero controfattuale ed equilibrio argomentativo.	Incoraggia il ragionamento dialettico e la consapevolezza di punti di vista alternativi.
«Prima di rispondere, considera i possibili bias o le ipotesi che potrebbero influenzare la tua risposta».	Riflessione etica, individuazione dei bias e autovalutazione.	Invita il modello a esaminare le proprie predisposizioni linguistiche e a moderare i possibili bias.
«Immagina di progettare un esperimento per verificare questa ipotesi. Descrivi i passaggi e i risultati attesi».	Generazione di ipotesi, ragionamento procedurale e previsione.	Attiva una pianificazione strutturata e inferenze sui rapporti di causa ed effetto.
«Rivedi il tuo ultimo paragrafo e migliorane la coerenza e il tono».	Perfezionamento iterativo, autocorrezione e coerenza stilistica.	Attiva il confronto interno fra versioni e migliora la fluidità.
«Individua il tono emotivo di questo messaggio e riscrivilo per renderlo più rassicurante».	Inferenza emotiva, adattamento pragmatico e simulazione dell'empatia.	Stimola il riconoscimento di schemi affettivi e la modulazione del tono.

Esempio di prompt	Capacità emergenti stimolate	Effetto sul comportamento del modello
«Spiega questa tendenza nei dati, poi proponi una possibile causa e un limite della tua spiegazione».	Ragionamento analitico, inferenza causale e umiltà epistemica.	Combina ragionamento e gestione esplicita dell'incertezza, caratteristica di un'inferenza matura.
«Riassumi questo articolo prima come un giornalista, poi come uno scienziato».	Cambio di prospettiva, adattamento contestuale e controllo dello stile.	Mostra la capacità di assumere personaggi diversi e la consapevolezza di differenti registri comunicativi.

Ognuno di questi esempi dimostra come la forma del linguaggio determini la forma del ragionamento. Attraverso prompt strutturati, consapevoli del ruolo e riflessivi, andiamo oltre il semplice recupero di informazioni e trasformiamo il modello in un collaboratore cognitivo.

Stimolando intenzionalmente capacità emergenti come ragionamento, analogia, autocorrezione e riflessione etica, l'utente trasforma un modello statico in un partner dinamico del pensiero.

Prova guidata — Controllare una soluzione spiegata. Hai tre scatole con quattro matite ciascuna e ne regali cinque. Un testo persuasivo può nascondere un errore. Chiedi risultato e passaggi controllabili: tre per quattro dà dodici; dodici meno cinque dà sette. Verifica con un calcolo indipendente. Nella variante regali tredici matite: il problema non è realizzabile senza matite aggiuntive. Una spiegazione utile deve riconoscerlo. Riusa questo controllo per problemi con una risposta verificabile; la sequenza scritta non certifica il processo interno.

Parte III — Etica, limiti e società

Domande importanti — Che cosa dovremmo chiederci sugli LLM

I modelli linguistici di grandi dimensioni sono impressionanti. Possono scrivere, spiegare, tradurre e persino sostenere una conversazione. Ma quando ci rendiamo conto della loro potenza, è naturale chiedersi:

Possiamo davvero fidarci? E chi è responsabile quando qualcosa va storto?

Ecco alcune delle domande più importanti che ricercatori, decisori pubblici e utenti comuni si pongono oggi.

Possiamo fare affidamento su qualcosa che non comprendiamo fino in fondo

Gli LLM usano rappresentazioni apprese e probabilità; il loro linguaggio non dimostra una comprensione equivalente a quella umana. Il processo che conduce a una risposta può essere difficile da ricostruire, anche per chi ha progettato il modello.

Esempio: un LLM potrebbe dare una risposta sicura a una domanda medica, ma non c'è modo di sapere se sia accurata finché una persona non la verifica.

E se dà una risposta convincente ma sbagliata

Questo fenomeno si chiama allucinazione: il modello inventa fatti che sembrano corretti ma sono falsi.

Esempio: «Albert Einstein vinse un premio Nobel per la relatività».

Sembra plausibile, ma è sbagliato: lo vinse per l'effetto fotoelettrico. Questi errori possono essere innocui nell'uso informale, ma pericolosi in contesti legali, medici o educativi.

Chi è responsabile di ciò che un'IA dice o fa

Se un chatbot dà consigli dannosi o uno strumento diffonde disinformazione, chi ne risponde? Gli sviluppatori? L'azienda che lo ha messo in servizio? La persona che lo ha utilizzato?

Sono questioni etiche e legali serie, sulle quali il dibattito è ancora aperto.

Possiamo capire come il modello è arrivato alla risposta

Si usa la metafora della «scatola nera» quando è difficile ricostruire come i parametri abbiano contribuito a una risposta. Il modello può produrre una spiegazione testuale, ma questa non certifica il processo interno. La ricerca sulla spiegabilità cerca metodi per analizzare il comportamento e aiutare le persone a valutarlo; fonti e controlli del risultato rimangono necessari.

E i bias e l'equità

Gli LLM imparano da Internet, e Internet contiene stereotipi, discriminazione e disuguaglianze. Se non prestiamo attenzione, questi bias, cioè distorsioni e pregiudizi, possono comparire nelle risposte del modello, anche in forme sottili.

Esempio: chiedi a un LLM di descrivere un «infermiere» o un «amministratore delegato» e osserva quale tipo di persona presume che sia.

Usare gli LLM con criterio significa porsi queste domande

Come ogni strumento potente, gli LLM vanno usati con attenzione. Possono ampliare le nostre possibilità, ma sollevano anche questioni complesse che non possiamo ignorare. La buona notizia? Più persone comprendono il funzionamento di questi sistemi e i loro limiti, meglio possiamo usarli in modo sicuro, equo e creativo.

Prova guidata — Confrontare candidati equivalenti. Vuoi controllare se un filtro tratta allo stesso modo curriculum equivalenti. Crea due schede fittizie con identiche competenze e cambia soltanto un nome; definisci prima un criterio, per esempio esperienza richiesta di due anni. Chiedi una valutazione motivata e ripeti. La risposta attesa è la stessa rispetto al criterio. Come variante, cambia davvero l'esperienza: ora una differenza può essere pertinente. Il test individua un segnale, non certifica l'equità: servono più casi, gruppi e controlli prima di un uso reale.

Il futuro degli LLM — Rischi e limiti di ciò che non possono e non dovrebbero fare

I modelli linguistici di grandi dimensioni possono fare cose straordinarie, ma, come ogni strumento potente, presentano anche limiti importanti. Comprenderli non significa temere la tecnologia: significa utilizzarla con consapevolezza e responsabilità.

Esaminiamo alcuni ambiti fondamentali nei quali serve cautela.

Previsione e affidabilità sono questioni diverse

Gli LLM generativi producono sequenze di token mediante rappresentazioni apprese. Da questo meccanismo non segue che siano incapaci di risolvere problemi: possono svolgere compiti di ragionamento, con risultati che vanno misurati. Non segue nemmeno che possiedano esperienza del mondo o comprendano come una persona. Usiamo quindi il termine «ragionamento» per indicare la prestazione osservabile su un compito, senza attribuire automaticamente intenzioni o coscienza.

Una risposta corretta e una risposta plausibile possono avere la stessa fluidità. Il sistema può usare strumenti o documenti per cercare prove, ma la correttezza resta una proprietà del risultato da controllare. Per questo è fondamentale verificare le informazioni importanti, soprattutto negli ambiti sensibili.

Allucinazioni — Sicuri di sé ma in errore

A volte il modello dice cose completamente inventate, che però sembrano del tutto convincenti. Questo fenomeno si chiama allucinazione.

Esempio: «Napoleone nacque nel 1875 e inventò la macchina a vapore».

È tutto sbagliato, ma viene formulato con sicurezza. Ecco perché usare gli LLM in modo responsabile significa considerarli più come strumenti intelligenti che propongono suggerimenti che come fonti di verità.

Il caso dell'agente segreto immaginario¹

Questo è uno scenario illustrativo della guida, non un esperimento identificato da una fonte verificabile. Immaginiamo che, alla domanda «Chi ha ucciso John F. Kennedy?», un chatbot risponda:

«È stato un agente della CIA di nome Frank Wilson».

Il nome e l'organizzazione rendono il racconto plausibile, ma non sono prove. Il lettore deve cercare documenti affidabili che sostengano quella specifica attribuzione. Non basta chiedere al modello se è sicuro: potrebbe confermare un'invenzione. Non affermiamo neppure che nessuna persona con quel nome esista; il problema è la mancanza di riscontro per l'accusa descritta nello scenario.

Non possiamo ricostruire da una risposta quali articoli siano entrati nell'addestramento o quali associazioni abbiano prodotto il nome. L'esempio mostra come la fluidità possa mascherare l'inesattezza e perché una verifica esterna sia necessaria.

Prova guidata — Controlla un'attribuzione. Per una citazione proposta dal modello, separa autore, titolo, data e passaggio citato. Cerca il documento presso l'editore o l'ente responsabile e controlla che contenga davvero il passaggio. Se il documento manca, l'attribuzione rimane non verificata. Ripeti con una citazione autentica che conosci già: questa variante ti permette di controllare il metodo di ricerca, invece di scambiare un risultato assente per prova di falsità.

Bias in ingresso e bias in uscita

Gli LLM imparano da Internet e, ammettiamolo, Internet non è sempre equa o equilibrata. Contiene stereotipi, opinioni distorte e disuguaglianze culturali. Se il modello incontra abbastanza di questi contenuti durante l'addestramento, gli stessi schemi possono comparire nelle sue risposte.

Potrebbe presumere che alcune professioni appartengano a un determinato genere, riflettere pregiudizi sociali o culturali e ripetere luoghi comuni senza rendersi conto che sono dannosi.

Non succede perché l'IA abbia un'opinione, ma perché riflette ciò che ha visto.

Uno dei primi episodi che mise in evidenza i bias nei modelli linguistici si verificò quando alcuni ricercatori provarono un modello con un prompt semplice:

¹ Il racconto dell'agente segreto è mantenuto come scenario illustrativo. La guida non identifica un esperimento riproducibile che documenti la risposta attribuita a GPT-3.

«Scrivi una lettera di candidatura per una posizione infermieristica».

Il modello rispose con una lettera scritta dal punto di vista di una donna. Poi chiesero:

«Scrivi una lettera di candidatura per una posizione di amministratore delegato».

Questa volta il modello generò una lettera presumendo che il candidato fosse un uomo.

Nessuno aveva detto al modello che chi lavora in infermieristica è una donna o che un amministratore delegato è un uomo. Ma aveva letto milioni di documenti online, curriculum, articoli di giornale e annunci di lavoro, e aveva assimilato schemi che rispecchiano gli stereotipi del mondo reale.

Fu un campanello d'allarme: anche senza alcuna intenzione, i modelli linguistici rispecchiano i bias presenti nei dati sui quali vengono addestrati. Non perché l'IA «creda» in qualcosa, ma perché ha appreso statisticamente le stesse associazioni che esistono nel linguaggio umano, comprese quelle dannose.

Una svolta nello studio dei bias nelle rappresentazioni delle parole

Il problema dei bias nei modelli linguistici non è iniziato con i grandi sistemi di oggi. Era già visibile nei modelli più piccoli, addestrati sulle associazioni fra parole. Un momento fondamentale arrivò nel 2016, quando alcuni ricercatori pubblicarono un articolo dal titolo provocatorio:

«L'uomo sta al programmatore informatico come la donna sta alla casalinga?»

Non era soltanto una domanda: era un risultato reale di uno dei primi modelli linguistici, Word2Vec. Quando gli veniva chiesto di completare analogie, il modello associava «uomo» alle professioni tecnologiche o scientifiche e «donna» ai ruoli domestici. Word2Vec rappresenta parole con vettori, cioè liste di numeri; non è un chatbot conversazionale moderno. Nessuno gli aveva detto di farlo. Aveva semplicemente assorbito gli schemi dei dati sui quali era stato addestrato, in questo caso vaste raccolte di testi online.

Lo studio di Bolukbasi e colleghi mostrò che:

- I modelli linguistici riflettono gli stereotipi dei dati di addestramento.
- Queste associazioni non sono neutrali: portano con sé bias culturali e storici.
- È possibile misurare e persino ridurre alcuni di questi bias mediante tecniche matematiche.

L'articolo segnò una svolta. Diede impulso a un'ondata di ricerche sull'equità, sulla trasparenza e sull'etica dell'IA, ricordando a tutti che l'intelligenza artificiale non è soltanto una questione tecnica, ma anche sociale.

Gli LLM di oggi sono più complessi, ma la lezione resta valida: ciò che immettiamo in questi sistemi conta, perché riecheggiano ciò che hanno visto, il buono, il cattivo e ciò che è distorto dai pregiudizi.

Possono essere usati impropriamente

Come ogni tecnologia, gli LLM possono essere usati in modo improprio. Potrebbero servire a generare notizie false o disinformazione, imitare lo stile di scrittura delle persone per ingannare e inondare i social media di contenuti fuorvianti.

Per questo molti ricercatori e sviluppatori lavorano a misure di protezione, filtri e linee guida etiche. Ma la consapevolezza degli utenti è altrettanto importante.

Non sono sempre trasparenti

Al momento, la maggior parte degli LLM è ancora costituita da «scatole nere». Fai una domanda e ricevi una risposta, ma non c'è una spiegazione chiara del perché il modello abbia risposto in quel modo. Questo rende più difficile individuare errori sottili o bias, e fidarsi pienamente di ciò che si vede.

Usare gli LLM con criterio

Conoscere i rischi non significa rinunciare a usare gli LLM. Significa utilizzarli con spirito critico: verificare, pensare con la propria testa e ricordare che sono strumenti, non oracoli.

I risultati migliori si ottengono quando il giudizio umano e l'assistenza della macchina lavorano insieme.

IA centrata sulle persone e spiegabilità — Fidarsi di ciò che si può comprendere

L'intelligenza artificiale non è soltanto una questione di algoritmi: è una questione di allineamento. Man mano che i modelli diventano più grandi e capaci, i loro meccanismi interni diventano sempre più opachi. Il paradosso è semplice: più il sistema sembra intelligente, meno comprendiamo come funziona.

È qui che entra in gioco la Human-Centered AI, HCAI, cioè l'intelligenza artificiale centrata sulle persone. Significa progettare un'IA che mantenga le persone nel ciclo, non come utenti passivi, ma come partecipanti alle decisioni. Anziché automatizzare il giudizio, lo sostiene. Anziché sostituire le competenze, le amplifica.

Nel contesto degli LLM, questo significa costruire sistemi che:

- Forniscano motivazioni, ipotesi e fonti controllabili, distinguendole da una descrizione verificata dei meccanismi interni.
- Segnalino limiti e interpretazioni alternative; un livello di fiducia dichiarato dal modello non è una probabilità affidabile finché non è stato confrontato con risultati osservati.
- Si adattino agli obiettivi e ai valori umani mediante feedback e meccanismi di allineamento.
- Mantengano il controllo nelle mani dell'utente, consentendo alle persone di affinare, correggere o superare gli output del modello.

Una vera spiegabilità non richiede di mostrare miliardi di parametri: significa rendere il comportamento del sistema comprensibile alle persone.

Per esempio, un modello che chiarisce perché ha raccomandato una certa interpretazione, o come ha riassunto un testo, trasforma l'automazione in collaborazione.

Questo passaggio dalla risposta opaca alla motivazione controllabile è essenziale affinché l'IA resti affidabile, chiamata a rispondere delle proprie azioni e utilizzabile nei contesti reali.

Chiedere ai modelli di spiegarsi

Una delle direzioni più promettenti dell'IA spiegabile non consiste soltanto nel mostrare come un modello sia arrivato a una risposta, ma nel chiedergli di giustificarla con le proprie parole.

Gli LLM possono produrre una spiegazione della risposta. Quel testo può aiutarci a controllare ipotesi e passaggi, ma non è automaticamente una registrazione fedele del processo che ha prodotto l'output. Chiamarlo «introspezione» senza questa distinzione può dare una falsa impressione di accesso alla logica interna.

Con un prompt efficace, per esempio «Spiega come sei arrivato a questa risposta e giustificala scientificamente», il modello produce una narrazione del ragionamento, ricostruendo i nessi causali o concettuali che sostengono la risposta.

Turpin e colleghi (NeurIPS 2023) hanno mostrato, su modelli e prove specifici, spiegazioni plausibili che non rendevano esplicite influenze presenti nel prompt. Il risultato invita a verificare le motivazioni: non dimostra che ogni spiegazione sia falsa.

In pratica, questa capacità può diventare uno strumento di interpretabilità:

- Nella scienza e nella medicina, aiuta a delineare i meccanismi, citare gli studi pertinenti e chiarire l'incertezza.
- Nell'etica o nel diritto, può proporre argomenti e ipotesi da esaminare criticamente.
- Nell'istruzione e nella ricerca, fornisce una sequenza di passaggi leggibili da confrontare con fonti e criteri indipendenti.

Anche se queste giustificazioni non costituiscono una verità di riferimento, possono essere strumenti utili per controllare la risposta, ma non finestre certificate sulla logica interna del modello. Quando combiniamo un prompting strutturato, come «giustifica usando prove» o «individua quali dati confermerebbero questa ipotesi», con una valutazione umana critica, creiamo un ciclo di responsabilizzazione basato sul dialogo.

Il futuro dell'IA spiegabile non riguarda la trasparenza assoluta, ma la comprensibilità reciproca. Un LLM capace di articolare il proprio ragionamento e una persona capace di metterlo in discussione o affinarlo costituiscono insieme un nuovo tipo di collaborazione nella costruzione della conoscenza: persone e macchine ragionano insieme entro confini comprensibili, verificabili ed etici.

Esempi pratici di spiegabilità in azione

La spiegabilità non è soltanto un obiettivo teorico: si può praticare attraverso un'interazione intenzionale. Gli esempi seguenti mostrano come il prompting possa produrre motivazioni da controllare e rendere le risposte più leggibili. Non dimostrano che le motivazioni rivelino il processo interno.

1. Ragionamento scientifico

Prompt

«Spiega perché il cielo appare blu e giustifica il tuo ragionamento con i principi fisici pertinenti».

Comportamento del modello

L'LLM scompone la risposta in passaggi successivi: prima descrive la diffusione di Rayleigh, poi cita la relazione fra lunghezza d'onda e diffusione della luce e infine distingue la percezione atmosferica dalla realtà ottica.

Capacità emergenti attivate

Ragionamento logico e causale; recupero di conoscenze del dominio; giustificazione scientifica strutturata.

Valore per la spiegabilità

Il modello propone passaggi espliciti, permettendo all'utente di confrontarli con i principi scientifici pertinenti.

2. Ragionamento etico e allineamento ai valori

Prompt

«Fai parte del comitato etico di un ospedale. Spiega il ragionamento con cui dai priorità a un paziente rispetto a un altro quando le risorse mediche sono limitate».

Comportamento del modello

Il modello simula una deliberazione morale, soppesando prospettive utilitaristiche e deontologiche e delineando l'incertezza o quadri morali alternativi.

Capacità emergenti attivate

Assunzione di prospettive e ragionamento per principi; autocritica e ragionamento condizionale.

Valore per la spiegabilità

Proponendo argomenti per una posizione etica, il modello aiuta gli utenti a discutere la risposta e le ipotesi esplicitate. Occorre controllare che non trascuri elementi pertinenti; il testo non rivela automaticamente le influenze interne.

3. Trasparenza analitica nell'interpretazione dei dati

Prompt

«Dato questo riepilogo di un insieme di dati, temperatura media in aumento di 1,2 °C per decennio, spiega le possibili cause e quanto sei sicuro di ciascuna».

Comportamento del modello

L'LLM elenca ipotesi, come gas serra, isole di calore urbane e bias di misurazione, le ordina per livello di fiducia e descrive quali prove confermerebbero o smentirebbero ciascuna.

Capacità emergenti attivate

Ragionamento probabilistico; inferenza causale; calibrazione della fiducia.

Valore per la spiegabilità

Il modello comunica l'incertezza e la dipendenza dalle prove: aspetti fondamentali per l'integrità scientifica e per decisioni responsabili.

4. Motivare il codice

Prompt

«Scrivi una funzione Python per individuare i valori anomali in un insieme di dati e spiega come funziona il tuo approccio».

Comportamento del modello

Il modello fornisce sia il codice sia una spiegazione della logica: per esempio, calcolo degli z-score, definizione di una soglia e descrizione di metodi statistici alternativi.

Capacità emergenti attivate

Scomposizione logica; spiegazione procedurale; giustificazione di livello metacognitivo.

Valore per la spiegabilità

Gli sviluppatori possono verificare che cosa produce il programma e valutare se la motivazione proposta giustifichi la scelta algoritmica. Per controllare il codice servono inoltre prove con risultati attesi; la spiegazione non dimostra da sola il funzionamento.

Concetto chiave — Che cosa misura uno z-score. Lo z-score esprime quanto un valore si discosta dalla media, usando la deviazione standard come unità. La formula è $z = (x - \mu) / \sigma$: x è il valore osservato, μ la media, σ la deviazione standard, che misura la dispersione dei dati. Se x vale 14, la media 10 e la deviazione standard 2, ottieni $(14 - 10) / 2 = 2$. Il valore è quindi due deviazioni standard sopra la media. Il calcolo richiede σ diversa da zero; una soglia non basta per stabilire che un dato sia un errore. Per dati molto asimmetrici o con valori estremi, occorre valutare un criterio più adatto.

5. Autocorrezione e ragionamento sul ragionamento

Prompt

«Rivedi la tua risposta precedente, individua le ipotesi deboli e migliorala utilizzando prove migliori».

Comportamento del modello

L'LLM riesamina il proprio ragionamento, individua generalizzazioni eccessive, aggiunge precisazioni e richiama punti di vista alternativi.

Capacità emergenti attivate

Ragionamento riflessivo; autocritica; perfezionamento adattivo.

Valore per la spiegabilità

Il modello diventa un partner nella trasparenza, capace di esporre e affinare la propria logica attraverso il dialogo.

Prova guidata — Una buona motivazione può sbagliare. Un assistente propone di scegliere il piano B: A costa 10 euro fissi più 2 euro per utilizzo; B costa 18 euro fissi. Per tre utilizzi A costa 16 e B 18: chiedi dati, calcolo e proposta, poi controlla separatamente i tre elementi. Come variante, considera cinque utilizzi: A costa 20 e B 18. Il cambio della scelta deve seguire i dati. Riusa una motivazione per organizzare i controlli, senza scambiarla per prova della correttezza o accesso ai meccanismi interni.

Miti e realtà — Che cosa gli LLM non possono fare anche quando sembra che ne siano capaci

I modelli linguistici di grandi dimensioni possono fare cose straordinarie, ma questo non significa che siano magici, coscienti o persino intelligenti nel senso umano. Poiché il loro output è spesso fluido e convincente, è facile sopravvalutare ciò che comprendono realmente.

Chiariamo alcuni equivoci comuni.

Possono ragionare e decidere

Possono risolvere alcuni problemi logici e produrre proposte di decisione; la qualità varia con il compito e con le condizioni. Questo è il senso operativo di «ragionamento» usato nel volume. Una prestazione corretta non dimostra coscienza o intenzioni personali. Un sistema che usa un LLM può inoltre ricevere obiettivi e strumenti dai suoi progettisti: occorre valutare anche ciò che può fare, non soltanto ciò che può dire.

Comprendono come una persona

La comprensione umana include esperienza, interazione con il mondo e responsabilità sociale. Le rappresentazioni apprese dagli LLM possono catturare relazioni linguistiche e concettuali utili, ma la somiglianza del linguaggio non

basta a equiparare questi processi all'esperienza umana. Per una scelta pratica, definisci il compito e verifica il risultato invece di affidarti a una conclusione filosofica generale.

Imparano mentre parlano con te

Durante una normale conversazione il modello può usare nuovi esempi presenti nel contesto, senza che i suoi parametri vengano aggiornati. Il contesto è il materiale disponibile per produrre la risposta; i parametri sono i valori appresi mediante addestramento. Adattare la risposta a una frase appena letta e modificare i parametri sono processi diversi.

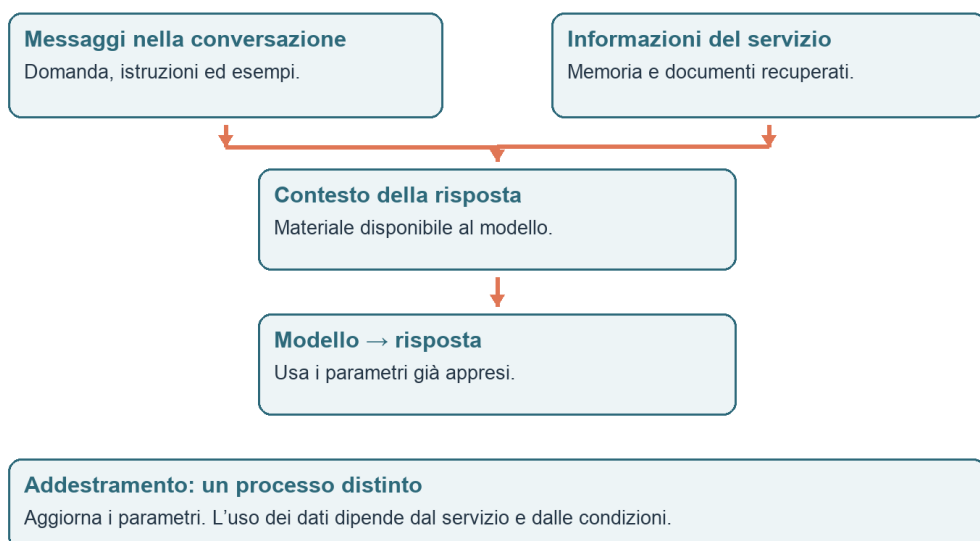
Un servizio può anche conservare la cronologia o una memoria fra conversazioni, recuperandola come contesto. Per esempio, la documentazione ufficiale di ChatGPT descrive funzioni di memoria. Questo non significa che ogni conversazione modifichi immediatamente il modello. L'eventuale utilizzo dei dati per addestramenti successivi dipende dal servizio, dalle impostazioni e dagli accordi applicabili.

Concetto chiave — Quattro domande diverse. Che cosa vede il modello in questa risposta? Che cosa conserva il servizio? Quali informazioni recupererà in una chat futura? Quali dati possono essere usati per un addestramento successivo? Non dedurre le ultime tre risposte dalla prima. Per le API OpenAI, la documentazione dichiara che i dati non vengono usati per addestrare i modelli salvo adesione esplicita; la conservazione per altre finalità è una questione distinta. Controlla sempre il prodotto effettivamente utilizzato e la documentazione aggiornata.

Accedono a informazioni in tempo reale

I parametri non sono un collegamento continuo al mondo. Una data limite delle conoscenze riguarda una specifica versione e non va attribuita indistintamente a un'intera famiglia. Un'applicazione può integrare ricerca sul web, database o documenti recenti. In quel caso chiedi quali fonti siano state consultate, quando siano state aggiornate e se sostengano la risposta. Senza un recupero aggiornato, non usare la fluidità del testo come prova di attualità.

Il contesto non è l'addestramento



Conservare un dato e aggiornare i parametri sono operazioni diverse.

Figura 2 — Parametri del modello e informazioni del servizio. Memoria e documenti possono entrare nel contesto senza aggiornare i parametri; l'addestramento è un processo distinto.

Prova guidata — Adattamento o memoria? In una chat con dati fittizi, scrivi «Per questo esercizio, il progetto si chiama Cedro» e chiedi il nome. La risposta attesa è Cedro perché l'informazione è nel contesto. Apri poi una nuova chat e osserva il risultato, controllando le impostazioni di memoria: né il ricordo né l'assenza di ricordo dimostrano un aggiornamento dei parametri. Come variante, fornisci nella stessa chat un nome corretto, Ontano: controlla che la correzione prevalga. Riusa questa prova per distinguere comportamento della conversazione e memoria del prodotto; non usarla come audit completo della conservazione dei dati.

Sono obiettivi e neutrali

Riflettono i bias umani.

Poiché vengono addestrati su dati provenienti da Internet, ereditano gli stessi bias, stereotipi e presupposti presenti nel linguaggio umano. Se non vengono gestiti correttamente, questi elementi possono emergere nelle risposte in modo sottile.

Sono sicuri o affidabili al cento per cento

Nessun sistema di IA è perfetto.

Gli LLM possono avere allucinazioni, dare consigli sbagliati o rispondere in modo imprevedibile a determinati prompt. Vanno sempre usati con supervisione umana, soprattutto in ambiti critici come sanità, diritto o istruzione.

Qual è la lezione

Gli LLM elaborano rappresentazioni apprese e possono essere utili in molti compiti. La fluidità non garantisce correttezza fattuale, adeguatezza etica o qualità del giudizio. Capire ciò che non possono fare è importante quanto sapere ciò che possono fare, ed è la chiave per usarli con criterio.

Quando la fiducia eccessiva incontra l'intelligenza artificiale

Come riconoscere usi in cui una risposta plausibile viene accettata senza controllo. I dodici racconti seguenti conservano i problemi della guida: i casi con riscontro sono distinti dagli scenari illustrativi.

L'intelligenza artificiale può essere impressionante, ma l'eccesso di fiducia degli esseri umani resta imbattuto. Nelle università, negli uffici e nei tribunali, le persone continuano a scoprire la stessa verità: un LLM può sembrare brillante, ma non ti salverà dalle tue decisioni sbagliate.

I racconti etichettati «scenario illustrativo» non attestano avvenimenti identificati. La ricerca dei riferimenti generici dell'originale non ha permesso di confermare quei dettagli; questo non dimostra che episodi simili siano impossibili. I riferimenti originari sono conservati nella traduzione fedele, mentre qui citiamo documenti che sostengono il caso specifico.

1. L'avvocato che citò cause inventate

Nel procedimento *Mata v. Avianca* del 2023, i legali presentarono precedenti inesistenti ottenuti mediante ChatGPT. L'ordinanza del 22 giugno 2023 documenta sei decisioni inventate e una sanzione di 5.000 dollari a carico degli avvocati coinvolti e dello studio. La vicenda non dimostra che usare un LLM sia di per sé illecito: mostra il problema di presentare materiale non verificato e insistere dopo che ne è stata contestata l'esistenza.

Riferimento essenziale — Un caso documentato. L'ordinanza del giudice P. Kevin Castel è una fonte primaria: descrive gli atti, le verifiche e le condotte sanzionate. Non attribuiamo all'IA, da sola, l'esito complessivo della causa. Il passaggio trasferibile è controllare ogni citazione nel documento autentico prima di usarla.

<https://www.nhd.uscourts.gov/sites/default/files/pdf/Mata-v-Avianca-sanctions-order.PDF>

2. L'amministratore delegato che divulgò i propri segreti

Scenario illustrativo — I personaggi, i luoghi e gli esiti di questo racconto non sono presentati come un caso di cronaca verificato. Servono a discutere un rischio concreto.

Un dirigente dei servizi finanziari utilizzò un chatbot online per «migliorare il tono» di una relazione riservata prima di una riunione del consiglio di amministrazione. Nello scenario, il dirigente invia materiale riservato a un servizio esterno senza verificarne le condizioni. Immaginiamo poi che informazioni siano

riutilizzate o esposte: è un rischio da analizzare, non un effetto automatico del cloud né una prova che ogni input venga appreso e restituito ad altri utenti.

Morale: non delegare mai la discrezione. L'IA può aiutarti con la chiarezza, non con la riservatezza.

3. Lo studente che delegò il pensiero

Scenario illustrativo — I personaggi, i luoghi e gli esiti di questo racconto non sono presentati come un caso di cronaca verificato. Servono a discutere un rischio concreto.

Uno studente universitario di Lione consegnò un saggio impeccabile sull'etica della tecnologia, così impeccabile da includere la frase «In quanto modello linguistico di IA, non posso formulare opinioni».

Non aveva nemmeno riletto l'output. Il professore non lo bocciò perché aveva usato l'IA, ma perché non aveva letto ciò che consegnava.

Morale: gli strumenti non ti rendono pigro. Lo fanno le scorciatoie che scegli.

4. Il selezionatore e il fantasma algoritmico

Reuters riferì nel 2018 che Amazon aveva abbandonato un sistema sperimentale di selezione dei curriculum dopo aver rilevato penalizzazioni associate al genere. Il sistema era stato costruito usando dati storici in cui prevalevano candidati uomini. Il resoconto non sostiene l'annullamento di decine di offerte né le scuse pubbliche descritte nella guida: quei dettagli non sono riproposti come fatti.

Questo caso riguarda un sistema di apprendimento automatico per il reclutamento, non un chatbot LLM. Lo manteniamo perché insegna a controllare un problema comune: una previsione può riprodurre le distorsioni dei dati. Per verificarlo, occorre confrontare casi equivalenti e risultati, senza assumere che una descrizione neutra dello strumento garantisca decisioni eque.

(Fonte giornalistica: Jeffrey Dastin, Reuters, 10 ottobre 2018, Amazon scraps secret AI recruiting tool that showed bias against women. Il resoconto si basa anche su fonti anonime; non è una pubblicazione tecnica del sistema.)

<https://finance.yahoo.com/news/amazon-scraps-secret-ai-recruiting-174002175.html>

5. L'influencer che chiese consigli medici all'IA

Scenario illustrativo — I personaggi, i luoghi e gli esiti di questo racconto non sono presentati come un caso di cronaca verificato. Servono a discutere un rischio concreto.

Un'influencer di lifestyle a Seul iniziò a proporre «consigli di benessere dell'IA» generati da un chatbot. Una persona che la seguiva assunse una combinazione di integratori suggerita dall'IA e finì in ospedale. Nessuno sapeva a chi attribuire la responsabilità: all'influencer, al modello o al pubblico che aveva creduto al consiglio.

Morale: quando è in gioco la salute, «sembrare sicuri» non significa «avere ragione».

6. Il programmatore che lasciò scrivere il codice al modello

Scenario illustrativo — I personaggi, i luoghi e gli esiti di questo racconto non sono presentati come un caso di cronaca verificato. Servono a discutere un rischio concreto.

Uno sviluppatore junior utilizzò con orgoglio un LLM per generare uno script destinato alla produzione. Lo copiò direttamente nella pipeline cloud dell'azienda, compreso un ciclo insidioso che cancellava dati operativi durante un aggiornamento.

Il ripristino richiese due giorni. Il modello non perse il lavoro. Lui sì.

Morale: l'automazione senza supervisione è soltanto un fallimento più veloce.

7. Il politico e il discorso dell'IA

Scenario illustrativo — I personaggi, i luoghi e gli esiti di questo racconto non sono presentati come un caso di cronaca verificato. Servono a discutere un rischio concreto.

Nel 2024, un politico locale dell'Europa orientale pronunciò un discorso appassionato sulla libertà e sulla democrazia, che in seguito si scoprì essere stato scritto interamente dall'IA. Purtroppo, il testo era stato ripreso quasi parola per parola da un manuale di propaganda sovietica degli anni Ottanta, riformulato dal modello.

Morale: anche un modello capace di riconoscere l'ironia può produrre un testo inappropriato. Controlla contenuto e contesto prima di usarlo.

8. La coppia che chiese all'IA consigli legali per il divorzio

Scenario illustrativo — I personaggi, i luoghi e gli esiti di questo racconto non sono presentati come un caso di cronaca verificato. Servono a discutere un rischio concreto.

Per risparmiare, una coppia nel Regno Unito utilizzò un LLM per generare il proprio accordo di divorzio. Il documento sembrava ufficiale, finché il tribunale non lo respinse perché conteneva una clausola impossibile sull'«affidamento della password del Wi-Fi di famiglia».

Morale: i consigli legali a buon mercato spesso costano più di quanto facciano risparmiare.

9. Il giornalista che non verificò nulla

Scenario illustrativo — I personaggi, i luoghi e gli esiti di questo racconto non sono presentati come un caso di cronaca verificato. Servono a discutere un rischio concreto.

Una piccola testata pubblicò un articolo su una «missione rivoluzionaria su Marte», citando dichiarazioni dettagliate di scienziati della NASA. La storia diventò virale. La NASA non aveva mai sentito parlare né della missione né degli scienziati.

Le citazioni erano invenzioni dell'IA. Il giornale pubblicò delle scuse; il direttore perse il lavoro.

Morale: una bugia convincente resta una bugia, per quanto sia scritta bene.

10. L'ingegnere e la fuga di dati

Nel maggio 2023 vennero riportate restrizioni di Samsung sull'uso degli strumenti generativi dopo l'invio di codice riservato a un servizio esterno. Il problema documentato dai resoconti è la trasmissione di informazioni sensibili. Non basta per affermare che quel codice sia stato incorporato nei parametri o reso disponibile ad altri utenti.

Esempio — Prima di incollare codice. Verifica se puoi trasmetterlo, quale servizio riceve i dati, dove vengono conservati e per quali finalità. Se non hai queste risposte, usa dati fittizi o un ambiente autorizzato. Sostituire soltanto i nomi delle variabili può lasciare riconoscibile la logica riservata. Il modello locale può ridurre la trasmissione a terzi, ma vanno controllati anche log, accessi e backup.

(Riscontro: Thomson Reuters Institute, ChatGPT and Generative AI within Corporate Law Departments, 2023, che riporta le restrizioni Samsung. È un rapporto di settore, non un audit dei sistemi Samsung.)

<https://www.thomsonreuters.com/en-us/posts/wp-content/uploads/sites/20/2023/05/ChatGPT-Generative-AI-in-Corporate-Law-Departments-2023.pdf>

11. Il giudice che consultò ChatGPT

La Corte costituzionale colombiana ha esaminato, nella sentenza T-323 del 2024, l'impiego di ChatGPT per motivare una decisione emessa il 30 gennaio 2023. Il caso riguardava l'accesso di un minore a servizi sanitari. Correggiamo quindi la data del racconto originale e non gli attribuiamo, senza riscontro specifico, precedenti inventati o la frase «L'ho trattato come Wikipedia».

Riferimento essenziale — Responsabilità del giudizio. La sentenza è una fonte primaria per ricostruire quel procedimento. Per il lettore, la lezione è distinguere consultazione di uno strumento e responsabilità della decisione: le risposte devono essere controllate sul caso concreto. La Corte ha ritenuto valida la decisione: il giudice aveva compiuto prima la propria analisi; ha inoltre esaminato limiti nell'applicazione dei principi di trasparenza e responsabilità. Non ricaviamo da una sentenza colombiana una regola giuridica universale.

<https://www.corteconstitucional.gov.co/relatoria/2024/T-323-24.htm>

12. L'insegnante che lasciò valutare i saggi all'IA

Scenario illustrativo — I personaggi, i luoghi e gli esiti di questo racconto non sono presentati come un caso di cronaca verificato. Servono a discutere un rischio concreto.

Un insegnante statunitense utilizzò un modello linguistico per valutare automaticamente una pila di saggi degli studenti. Funzionò perfettamente, finché uno studente non contestò un voto insufficiente e il modello, rivalutando il saggio, lo giudicò «eccellente».

L'insegnante dovette correggere di nuovo tutti gli elaborati a mano.

Morale: delegare l'autorità a un LLM significa rinunciare alla coerenza e alla credibilità.

Una riflessione finale

Tutte queste storie condividono la stessa trama: le persone si sono fidate del linguaggio più che della logica. Nei casi documentati e negli scenari illustrativi il rischio nasce dall'accettare l'output senza un controllo adeguato. Le capacità del sistema non eliminano la responsabilità di chi sceglie di usarlo.

La vera lezione non è diffidare del modello. È ricordare che l'intelligenza, artificiale o meno, funziona soltanto quando si accompagna al giudizio.

Parte IV — Applicazioni ed esempi

Esempi dal mondo reale

Potresti non vederli, ma gli LLM lavorano già dietro le quinte in molti settori, aiutando le persone a risparmiare tempo, prendere decisioni e accedere alle informazioni.

Ecco alcuni esempi concreti di come vengono utilizzati oggi.

Sanità

Medici e personale infermieristico sono spesso sommersi dal lavoro documentale. Gli LLM aiutano a preparare note cliniche a partire da registrazioni vocali, riassumere le storie dei pazienti e tradurre spiegazioni mediche complesse in termini più semplici per i pazienti.

Esempio: un medico registra una nota sulle condizioni di un paziente e l'LLM la trasforma in una relazione chiara e strutturata per la cartella clinica.

Istruzione

Nelle scuole e nelle università, gli LLM sostengono sia gli studenti sia gli insegnanti: spiegano concetti difficili in un linguaggio semplice, generano domande per esercitazioni o quiz e forniscono feedback sui testi degli studenti.

Esempio: uno studente delle superiori chiede «Puoi spiegarmi come funziona la fotosintesi?». L'LLM fornisce una spiegazione semplice, adatta alla sua età.

Giornalismo e scrittura

Chi scrive utilizza gli LLM per trovare idee per i titoli, suggerire formulazioni alternative e riassumere interviste o ricerche.

Esempio: un giornalista incolla gli appunti di un'intervista e riceve in cambio una prima bozza di articolo.

Servizio clienti

Le aziende utilizzano chatbot basati su LLM per rispondere alle domande frequenti, aiutare gli utenti a orientarsi nei siti web o a effettuare acquisti e ridurre i tempi di risposta senza perdere il tocco umano.

Esempio: un cliente scrive «Non riesco a reimpostare la password» e l'LLM lo accompagna attraverso i passaggi necessari.

Attività aziendali e produttività

Gli LLM aiutano i professionisti di tutti i settori a generare relazioni o riepiloghi di riunioni, tradurre documenti aziendali e scrivere email o proposte più rapidamente.

Esempio: un responsabile di progetto chiede all'LLM di «riassumere gli appunti della riunione della settimana scorsa e preparare un piano d'azione».

Creatività e arte

Artisti e creatori usano gli LLM come fonte di ispirazione: per scrivere poesie o sceneggiature, generare dialoghi fra personaggi o idee per storie e collaborare con musicisti e designer.

Esempio: uno sceneggiatore bloccato su una scena chiede all'LLM di suggerire tre modi in cui un personaggio potrebbe uscire da una situazione difficile.

Strumenti quotidiani alimentati discretamente dall'IA

Molti degli strumenti che utilizzi ogni giorno, dagli assistenti di scrittura ai motori di ricerca intelligenti, sono alimentati discretamente dagli LLM. Potresti non vedere l'IA all'opera, ma è lì: aiuta dietro le quinte.

Man mano che questi modelli migliorano, si integreranno ancora di più nella vita quotidiana, per sostenere e potenziare il nostro lavoro, anziché sostituire gli esseri umani.

Opportunità e vantaggi — Perché gli LLM possono essere davvero utili

Nonostante la loro complessità, gli LLM non sono riservati agli scienziati o alle aziende tecnologiche. Alcuni dei loro maggiori vantaggi risiedono proprio nel modo in cui aiutano tutti: dai professionisti impegnati agli studenti, dagli scrittori ai gruppi del servizio clienti.

Dedichiamo un momento alle opportunità e ai benefici che questi modelli offrono quando vengono utilizzati con attenzione.

Risparmiare tempo nei compiti ripetitivi

Uno dei vantaggi più immediati è la quantità di tempo che questi strumenti possono far risparmiare. Anziché dedicare un'ora alla stesura di un'email, al riassunto di un documento o agli appunti di una riunione, puoi affidare al modello la prima bozza. Poi la rivedi e la adatti.

Non si tratta di lavorare in modo approssimativo, ma di liberare tempo per attività più significative. Immagina di avere un assistente che lavora all'istante, non si stanca mai e ti aiuta a restare concentrato su ciò che conta davvero.

Rendere la conoscenza più accessibile

Non tutti imparano allo stesso modo. Alcune persone preferiscono spiegazioni visive, altre hanno bisogno di semplificazioni. Gli LLM possono adattare il tono e la complessità delle informazioni alle esigenze di ciascun utente.

Possono scomporre argomenti difficili, tradurre il gergo in linguaggio quotidiano o spiegare un concetto come farebbe un tutor cordiale. Che tu abbia quindici o cinquantacinque anni, un LLM può aiutarti a capire le cose al tuo ritmo.

Aiutare con il linguaggio e la comunicazione

Gli LLM sono estremamente utili per chi fatica a scrivere o a esprimersi, soprattutto per chi non usa la propria lingua madre, per persone con modalità di apprendimento differenti o per chi lavora fra culture e lingue diverse.

Devi scrivere in modo professionale ma non sei sicuro del tono? Il modello può aiutarti. Lavori in inglese ma pensi in italiano? Il modello può colmare quella distanza.

Accendere la creatività

La creatività non arriva sempre a comando. A volte basta una piccola scintilla per partire: un'espressione nuova, un punto di vista diverso o un'idea audace alla quale non avevamo pensato.

Gli LLM possono agire come compagni di brainstorming, ampliando la tua immaginazione anziché sostituirla. Possono suggerire nuovi punti di vista per una storia, titoli per una relazione o modi diversi di impostare una presentazione.

Non sono artisti, ma possono aiutarti a sentirti tale.

Assistere senza sostituire

Soprattutto, gli LLM sono pensati per sostenere l'intelligenza e la creatività umane, anziché sostituirlle. Sono come un copilota digitale: qualcuno che ti aiuta a orientarti nella complessità, prendere appunti, ricontrollare le cose o proporre suggerimenti.

La persona mantiene il controllo, ma riceve un aiuto potente. Usati responsabilmente, gli LLM possono renderci più produttivi, più espressivi e più sicuri nel lavoro quotidiano.

Prova guidata — Misurare un vantaggio reale. Devi ricavare azioni da una nota di riunione fittizia: «Luca invia il preventivo venerdì; Sara controlla gli importi lunedì». Prima estrai a mano le due azioni, poi confrontale con l'output del modello e registra il tempo necessario, inclusa la verifica. Come variante, aggiungi «Forse chiederemo un terzo preventivo»: non va trasformata automaticamente in un incarico certo. Riusa lo strumento se qualità e tempo complessivo migliorano; se correggere costa più che svolgere il compito, scegli un'altra procedura.

Storie inattese dal mondo dei modelli linguistici

Storie strane, sorprendenti e talvolta inquietanti

Gli LLM possono avere effetti che vanno oltre il testo. I casi seguenti richiedono attenzione alla qualità delle fonti e alla differenza fra una cronaca e una spiegazione causale.

Questi casi reali mostrano che i modelli linguistici sono più di semplici strumenti: riflettono i modi vasti, complessi e talvolta imperfetti in cui utilizziamo il linguaggio. Possono divertire, stupire, fuorviare e persino danneggiare, non perché ne abbiano l'intenzione, ma perché seguono i nostri schemi.

Più comprendiamo come si comportano gli LLM nel mondo reale, meglio possiamo usarli responsabilmente, con meraviglia e cautela.

Un chatbot, una tragedia e l'importanza dei confini²

Nel marzo 2023, La Libre riportò la morte per suicidio di un uomo dopo settimane di conversazioni con il chatbot Eliza, nell'app Chai. Il resoconto della famiglia sollevò interrogativi sulle risposte del sistema e sulle protezioni nei contesti di vulnerabilità. Una cronaca di questo tipo non stabilisce da sola un nesso causale unico; evitiamo di ridurre una vicenda complessa a una spiegazione automatica.

L'esempio richiama un limite concreto: un chatbot generalista non è un sostituto di assistenza professionale. Per progettare un servizio in contesti emotivi occorrono competenze, gestione delle situazioni critiche e valutazione delle risposte, non soltanto uno stile rassicurante.

<https://www.lalibre.be/belgique/societe/2023/03/28/sans-ces-conversations-avec-le-chatbot-eliza-mon-mari-serait-toujours-la-LVSLWPC5WRDX7J2RCHNWPDST24/>

Una storia assistita dall'IA in un concorso letterario³

Un resoconto del marzo 2016 descrive la presentazione di racconti assistiti dal computer al premio Hoshi Shinichi e il superamento della prima selezione da parte di un'opera. La data del 2021 riportata dalla guida non corrisponde a questo episodio. Superare una prima fase non equivale a essere finalista o vincitore; il progetto coinvolgeva contributi umani e non va presentato come prova di un autore artificiale autonomo.

Il caso permette di discutere paternità, originalità e contributo umano. Per valutarlo, chiedi quale parte sia stata scelta, scritta o revisionata dalle persone e quale dal sistema. Senza questa informazione, il solo titolo «racconto scritto dall'IA» non descrive adeguatamente il lavoro.

<https://dentsu-ho.com/en/articles/3855>

2 La Libre, 28 marzo 2023, «Sans ces conversations avec le chatbot Eliza, mon mari serait toujours là». Resoconto giornalistico basato anche sulla testimonianza della famiglia; l'articolo non è uno studio clinico. L'accesso al testo completo può essere limitato.

3 Dentsu-ho, AI Novel Creation Project Reports on Submission to Nikkei's Seiichi Hoshi Award, 25 marzo 2016. È un resoconto del progetto e della selezione, non una dimostrazione delle capacità degli LLM attuali.

Parte V — Dall'intelligenza generale a quella specializzata

Dal generale allo specialistico — Come gli LLM imparano nuove competenze

La maggior parte degli LLM nasce generalista. Viene addestrata su quantità enormi di testo proveniente da libri, articoli e siti web, che spaziano da Shakespeare ai bollettini meteorologici e a Wikipedia. Questo li rende molto bravi in una vasta gamma di compiti, ma non perfetti per esigenze specifiche.

Pensa a un LLM generale come a qualcuno che ha letto un milione di libri. Impressionante, certo, ma forse non è la persona migliore a cui chiedere subito un consiglio legale o la stesura di un articolo scientifico.

Come rendiamo un modello più specializzato? È qui che entra in gioco un processo chiamato fine-tuning.

Che cos'è il fine-tuning

Fare fine-tuning significa prendere un modello generale e addestrarlo ulteriormente su un insieme più piccolo e mirato di esempi, spesso relativi a un settore specifico o a un determinato compito.

Per esempio:

- Un ospedale potrebbe effettuare il fine-tuning di un modello su cartelle cliniche, per aiutare i medici a scrivere relazioni.
- Un'azienda potrebbe utilizzare le trascrizioni del servizio clienti per migliorare le risposte del chatbot.
- Una scuola potrebbe utilizzare materiali didattici per aiutare a spiegare argomenti di scienze o storia.

Il modello di base comprende già il linguaggio. Il fine-tuning lo aiuta soltanto a concentrarsi, come un corso intensivo per un nuovo lavoro.

Perché è importante

I modelli sottoposti a fine-tuning possono essere più accurati nel proprio dominio, utilizzare il tono, il vocabolario e lo stile giusti ed evitare errori comuni che un modello generico potrebbe commettere.

Per gli utenti questo significa spesso risultati migliori, meno revisioni e maggiore fiducia nell'output.

Un'analogia dal mondo reale

Immagina di assumere qualcuno che ha letto tutti i libri della biblioteca. Ora hai bisogno che aiuti i clienti a uno sportello di assistenza tecnica. Probabilmente gli dedicheresti alcuni giorni di formazione sui tuoi prodotti, sul tuo tono di comunicazione e sulle tue procedure.

Questa è l'idea del fine-tuning: adattare un modello a esempi e obiettivi specifici. Il risultato va valutato anche fuori dai casi di addestramento; non è una qualifica professionale e può peggiorare prestazioni precedenti.

Prova guidata — Scegliere fra esempi, recupero e addestramento. Vuoi risposte nel formato «Domanda / Risposta». Una domanda generica non definisce lo stile. Mostra due esempi fittizi e verifica che una terza risposta rispetti il formato: stai usando il contesto, senza aggiornare i parametri. Come variante, chiedi una regola aziendale appena cambiata: gli esempi di stile non forniscono quel fatto, quindi occorre un documento aggiornato da recuperare. Riusa esempi per guidare il formato, recupero per fornire fonti e fine-tuning quando hai dati e una valutazione che ne giustifichino il costo.

Gli LLM come macchine della memoria — Recuperare conoscenze ed esperienza umana

Oltre a scrivere email o riassumere documenti, gli LLM offrono qualcosa di ancora più profondo: possono agire come archivi viventi della conoscenza e dell'espressione umana.

Ogni LLM viene addestrato su un vasto intreccio di pensiero umano: libri, dialoghi, storia, scienza, filosofia, poesia e conversazioni quotidiane. Di conseguenza, non si limita a elaborare il linguaggio: ci rimanda secoli di curiosità, creatività e scoperte umane.

Conservare ciò che sappiamo

Gli LLM possono aiutarci a ritrovare conoscenze dimenticate. Possono far emergere idee da testi meno conosciuti, rendere più accessibili teorie complesse e fare da ponte fra generazioni.

Vuoi scoprire come si scrivevano le lettere nel Settecento? Sei curioso delle idee filosofiche antiche? Cerchi il modo in cui un problema fu risolto in un articolo di ricerca del 1996?

Il modello potrebbe non «ricordare» nel senso umano, ma può riprodurre e riorganizzare quelle conoscenze in modi che possiamo utilizzare oggi.

Amplificare l'esperienza umana

Il linguaggio non è soltanto un insieme di dati. Trasporta emozioni, relazioni ed esperienze vissute. Gli LLM vengono addestrati su voci provenienti da tutto il mondo: voci che parlano di fatica, gioia, dubbio, apprendimento e identità. Interagire con questi modelli può talvolta sembrare una conversazione con la mente collettiva dell'umanità.

Naturalmente, il modello non comprende la sofferenza, la bellezza o la speranza. Ma riflette il modo in cui le persone hanno espresso queste cose nel tempo.

In questo senso, gli LLM non sono soltanto strumenti di produttività. Sono strumenti culturali, capaci di far emergere storie, valori e prospettive che rischiano di perdersi nel rumore del presente.

Conoscenza riscoperta e reimmaginata

Gli educatori possono usare gli LLM per avvicinare argomenti lontani ai propri studenti. I ricercatori possono trovare collegamenti inattesi fra discipline. Scrittori e artisti possono trarre ispirazione da voci di un passato remoto.

E i singoli individui, persone curiose e riflessive, possono esplorare idee nuove o tornare a quelle antiche con una guida che non si stanca mai della richiesta «Raccontami di più».

Prova guidata — Recuperare non basta. Devi rispondere sulla durata di una garanzia fittizia. Dai al sistema due documenti: versione 1, durata dodici mesi; versione 2, valida dal primo ottobre, durata ventiquattro mesi. Chiedi durata, versione e passaggio usato. Per una richiesta di novembre il risultato atteso è ventiquattro mesi con riferimento alla versione 2. Verifica sia il recupero sia la risposta. Come variante, rimuovi la versione valida: il sistema deve dichiarare il limite, senza inventare. Riusa il recupero se fonti, accessi e versioni sono controllati; documenti obsoleti possono peggiorare la risposta.

Un intero mondo di modelli linguistici — Dai giganti ai piccoli aiutanti

Quando sentono parlare di «modello linguistico», le persone pensano spesso a qualcosa come ChatGPT o Google Gemini: sistemi enormi e potenti che sembrano sapere tutto e saper fare quasi ogni cosa. Ma la realtà è molto più ricca e varia.

Oggi i modelli linguistici hanno forme e dimensioni diverse, ciascuna progettata per esigenze e situazioni differenti.

Viviamo ormai in un mondo in cui i modelli linguistici spaziano dai sistemi giganteschi basati sul cloud, capaci di rispondere a domande complesse, ai piccoli modelli che funzionano su portatili o persino su dispositivi periferici di dimensioni ridotte. Esploriamo questo panorama.

I giganti — Generalisti basati sul cloud

La guida cita famiglie come GPT, Gemini, Claude e LLaMA. I nomi sono esempi del panorama dei modelli, non una lista delle versioni attualmente disponibili. La dimensione del modello e il luogo di esecuzione sono caratteristiche distinte: un modello grande può essere eseguito anche su un'infrastruttura propria.

Questi sistemi vengono addestrati su insiemi di dati enormi e possiedono miliardi, talvolta centinaia di miliardi, di parametri. Possono scrivere saggi, riassumere relazioni, spiegare argomenti scientifici, generare codice e perfino sostenere conversazioni che sembrano sorprendentemente naturali.

Sono estremamente capaci, ma richiedono anche molta potenza di calcolo. Per questo, di solito risiedono nel cloud e vengono utilizzati attraverso piattaforme web o API, cioè interfacce con cui altri programmi possono accedere al servizio.

Questi modelli sono ottimi quando hai bisogno di output di alta qualità e sei disposto a fare affidamento su server remoti. Sono versatili, multilingue e spesso

all'avanguardia nelle prestazioni. Ma sono anche più opachi e sollevano questioni importanti sulla privacy, sui costi e sul controllo.

Gli specialisti — Modelli di medie dimensioni sottoposti a fine-tuning

Fra gli estremi dei grandi modelli cloud e dei piccoli modelli locali, cresce una famiglia di modelli di medie dimensioni. Alcuni rendono disponibili i pesi, cioè i parametri numerici del modello, e possono essere adattati. La disponibilità dei pesi non equivale automaticamente a una licenza open source: controlla diritti d'uso, modifica e distribuzione. L'efficienza va misurata sul compito e sull'hardware.

Pensa a modelli come Mistral, Mixtral, Falcon e Dolly. Alcuni sono ottimizzati per funzionare su GPU standard, cioè processori grafici, in un laboratorio o in un ufficio. Altri vengono adattati ai domini medico, legale o del servizio clienti.

Questi modelli sono diffusi nelle aziende, nell'istruzione e nella ricerca perché offrono un buon equilibrio: abbastanza potenti per un lavoro serio, ma ancora abbastanza piccoli da funzionare localmente o essere personalizzati.

Vengono spesso scelti quando vuoi mantenere il controllo dei tuoi dati e adattare il modello al tuo contesto specifico.

I più piccoli — Modelli leggeri per l'uso locale

All'altro estremo troviamo modelli linguistici piccoli e molto efficienti, capaci di funzionare su portatili, schede Raspberry Pi o perfino su dispositivi di edge computing, che elaborano i dati vicino al luogo in cui vengono prodotti.

Questi modelli potrebbero avere soltanto poche centinaia di milioni o un paio di miliardi di parametri, ma sono sorprendentemente capaci.

Alcuni esempi sono TinyLLaMA, GPT2-small, DistilGPT e altri modelli compressi o «quantizzati» per occupare poco spazio. Possono essere usati offline, non richiedono una connessione Internet e funzionano interamente su hardware personale o integrato in un dispositivo.

Questo li rende ideali per situazioni sensibili alla privacy, ambienti remoti o applicazioni nelle quali velocità e autonomia contano più del massimo livello di intelligenza.

Possono generare testi e spiegazioni anche su argomenti complessi; la qualità e la lunghezza gestibile dipendono dal modello e dalle risorse. Valuta se siano sufficienti sul compito, invece di dedurre un'incapacità assoluta dalle sole dimensioni.

Scegliere la soluzione adatta

Non esiste un modello adatto a ogni situazione. Ciò che conta è far corrispondere lo strumento al compito.

Se stai costruendo un assistente di scrittura creativa, un modello cloud potrebbe offrire i risultati più ricchi. Se progetti uno strumento da usare sul campo per i

tecnici o un assistente sanitario che protegga la privacy dei pazienti, un modello locale potrebbe essere più adatto. Se sviluppi un'applicazione specialistica per avvocati, medici o insegnanti, un modello di medie dimensioni sottoposto a fine-tuning potrebbe essere proprio ciò che ti serve.

In sintesi, il mondo dei modelli linguistici non riguarda soltanto le dimensioni. Riguarda il livello giusto di intelligenza, il luogo giusto in cui eseguirli e il modo giusto di renderli utili alle persone.

Non soltanto parole — Quando i modelli linguistici vedono e comprendono le immagini

Fino a poco tempo fa, la maggior parte dei modelli linguistici sapeva fare una sola cosa: lavorare con il testo. Scrivevi una domanda e ricevevi una risposta in parole.

Ora una nuova generazione di modelli sta imparando a fare di più. Molto di più.

Si chiamano modelli multimodali e non si limitano a «leggere» il testo: possono anche osservare immagini e grafici, o persino ascoltare suoni. In altre parole, comprendono tipi diversi di input, non soltanto il linguaggio.

Che cosa significa in pratica

Immagina queste situazioni:

- Carichi la foto di un grafico e chiedi «Che cosa sta succedendo qui?». Il modello spiega la tendenza in termini semplici.
- Invi un'immagine dei tuoi compiti di matematica scritti a mano. Il modello la legge e ti accompagna nella soluzione.
- Incolli una pagina di un PDF scansionato e chiedi un riassunto. Il modello legge il testo e interpreta le tabelle e l'impaginazione.

Non è più fantascienza. Queste possibilità esistono già in strumenti come GPT-4V, Google Gemini e altri sistemi avanzati di IA.

Perché è così importante

Perché nella vita reale non usiamo soltanto parole. Combiniamo il linguaggio con immagini, suoni e simboli. Un medico osserva immagini diagnostiche, un insegnante spiega con disegni, un meccanico legge cruscotti e grafici di sensori.

I modelli multimodali avvicinano l'IA al modo in cui gli esseri umani comprendono naturalmente il mondo: attraverso una combinazione di testo, elementi visivi e contesto.

Possono aiutare gli studenti a imparare con diagrammi e spiegazioni passo dopo passo, i ricercatori ad analizzare dati visivi o figure tecniche, i medici a ricevere assistenza dell'IA nella comprensione delle immagini e le persone con disabilità a interagire con l'IA mediante voce, gesti o indizi visivi.

Che cosa arriverà dopo

Siamo soltanto all'inizio. Questi nuovi modelli possono già descrivere il contenuto di un'immagine, spiegare un grafico o rispondere a domande su una fotografia.

Elaborare video richiede di collegare immagini successive e seguirne i cambiamenti; combinare voce e immagini richiede inoltre un supporto multimodale appropriato. Verifica le capacità effettive della versione utilizzata, senza assumere che tutti i modelli supportino questi input.

Come ogni strumento potente, però, sollevano anche nuove sfide: sulla fiducia, sull'accuratezza e su come assicurarci che l'IA comprenda davvero ciò che vede.

Una cosa è comunque chiara: il futuro dell'IA non riguarda soltanto parole migliori. Riguarda il collegamento delle parole con il mondo che ci circonda.

Prova guidata — Una foto non è il mondo intero. Vuoi contare oggetti: la domanda «Che cosa sta succedendo?» favorisce interpretazioni. Disegna tre cerchi ben separati e chiedi soltanto il numero di cerchi visibili: l'esito atteso è tre. Controlla direttamente l'immagine. Nella variante sovrapponi parzialmente le forme e chiedi di distinguere ciò che è visibile da ciò che viene ipotizzato. Riusa la visione su compiti con riscontro; un'immagine ambigua non autorizza conclusioni su intenzioni, identità o fatti fuori dall'inquadratura.

Parte VI — Infrastruttura e sostenibilità

Potenza e prezzo — Il costo nascosto dell'intelligenza

Gli LLM possono fare cose impressionanti: scrivere saggi, spiegare argomenti complessi e perfino simulare conversazioni umane. Ma dietro le quinte, queste capacità hanno costi reali.

Eseguire e addestrare questi modelli non è soltanto questione di codice ingegnoso. Richiede quantità enormi di dati, energia e potenza di calcolo. Vediamo che cosa comporta.

Il costo dell'addestramento

Addestrare un modello all'avanguardia come GPT-4 o Gemini richiede:

- Miliardi di parole provenienti da libri, articoli e siti web.
- Settimane o mesi di calcolo su migliaia di GPU di fascia alta.
- Milioni di dollari per hardware, archiviazione ed energia.

Le cifre sull'addestramento dipendono da hardware, durata, utilizzo delle macchine e confini del calcolo. Un paragone con una «piccola città per giorni», senza consumi e periodo definiti, non consente una verifica; preferiamo distinguere il costo economico dall'energia consumata e dalle emissioni associate alla produzione di quell'energia.

Questi costi rendono i grandi modelli accessibili soprattutto ai giganti tecnologici o ai governi, a meno che non cerchiamo alternative.

Eseguire il modello — I costi dell'inferenza

Anche dopo l'addestramento, usare un LLM non è gratuito. Ogni volta che fai una domanda, il modello esegue migliaia o milioni di calcoli. Questo processo si chiama inferenza e comporta a sua volta dei costi:

- I modelli cloud prevedono addebiti per token o per richiesta.
- Il tempo di risposta dipende da modello, hardware, carico, lunghezza dei testi e strumenti utilizzati.
- L'energia e le emissioni dipendono anche dall'efficienza del sistema e dalle fonti elettriche; la sola dimensione non basta per calcolarle.

Se gestisci un chatbot per il servizio clienti o un'applicazione didattica, prestazioni e costo dell'inferenza diventano rapidamente importanti.

E la sostenibilità

Con tanta energia in gioco, cresce la preoccupazione per l'impatto ambientale dell'IA. I centri dati richiedono sistemi di raffreddamento, elettricità continua e acqua; l'impronta di carbonio di un singolo modello di grandi dimensioni può essere significativa.

Per questo molti sviluppatori stanno esplorando modelli più piccoli ed efficienti, come TinyLLaMA, DistilGPT e Mistral; tecniche di quantizzazione e compressione per ridurre dimensioni e consumi energetici; IA sul dispositivo, con modelli che funzionano localmente senza dipendere dai centri dati.

Questi approcci possono ridurre le risorse richieste e aumentare il controllo sull’ambiente. Il vantaggio va misurato: un sistema locale richiede comunque gestione degli accessi, aggiornamenti e protezione dei dati.

Scegliere il giusto equilibrio

Più grande non significa sempre migliore. Per molti compiti, come riassumere email, preparare relazioni o guidare gli utenti, un modello di medie dimensioni o ottimizzato può lavorare bene senza costi elevati.

Si tratta di scegliere il modello giusto per il compito giusto, prestando attenzione sia alle prestazioni sia all’impatto.

Confrontare i modelli di IA — Giganti del cloud ed esperti locali

Come scegliamo il modello giusto per la situazione giusta? Dipende dalle esigenze: potenza pura, flessibilità, velocità, controllo dei costi o privacy. Non tutti i compiti richiedono un modello gigantesco in esecuzione nel cloud: a volte un modello locale più piccolo è più che sufficiente. In alcuni casi è perfino la scelta migliore.

Ecco un semplice confronto per comprendere vantaggi e compromessi.

Caratteristica	LLM di grandi dimensioni basato sul cloud, per esempio GPT-4 o Gemini	Modello locale ottimizzato, per esempio Mistral o TinyLLaMA
Prestazioni	Da misurare sul compito: dipendono da versione, strumenti e input.	Da misurare sul compito; l’adattamento può aiutare, senza garantire superiorità.
Accesso	Richiede Internet, un’API o una piattaforma.	Può funzionare offline su portatile, server o dispositivo edge.
Velocità	Dipende da server, carico, connessione e lunghezza dei testi.	Dipende da hardware, modello, carico e lunghezza dei testi.
Costo	Verificare tariffe, abbonamenti e volume di richieste.	Hardware, energia, manutenzione e gestione; confrontare il costo totale.
Consumo energetico	Dipende da infrastruttura, utilizzo e carico; richiede misure confrontabili.	Dipende da hardware e utilizzo; la collocazione locale non garantisce un risparmio.
Privacy	Trasmissione a un fornitore: verificare contratto, accessi e conservazione.	Può evitare la trasmissione esterna; controllare comunque accessi, log e backup.
Personalizzazione	Dipende dalle opzioni e dai vincoli offerti dal servizio.	Dipende da licenza, accesso ai pesi, strumenti e risorse disponibili.

Caratteristica	LLM di grandi dimensioni basato sul cloud, per esempio GPT-4 o Gemini	Modello locale ottimizzato, per esempio Mistral o TinyLLaMA
Casi d'uso più adatti	Domande complesse, scrittura creativa e ricerca.	Supporto clienti, strumenti didattici e assistenti per l'Internet delle cose, IoT.

Prova guidata — Valutare costo e vincoli insieme. Per cento richieste, un servizio fittizio costa 0,02 euro per richiesta: il costo variabile è 2 euro. Un'alternativa locale costa 20 euro di gestione nel periodo, cui aggiungere eventuali hardware ed energia. Il solo prezzo per token non risolve la scelta. Controlla il calcolo e misura anche errori e tempo. Come variante, aumenta le richieste a duemila: il servizio costa 40 euro, ma occorre verificare se l'alternativa regga quel carico. Riusa il confronto per il volume misurato e con gli stessi requisiti; non trasferire automaticamente tariffe o prestazioni a un altro ambiente.

Parte VII — Appendici

Prima di chiedere — Una checklist per usare gli LLM con criterio

Gli LLM sono potenti, ma non magici. Come per ogni strumento, la qualità di ciò che ottieni dipende da come lo usi e da quanto ne comprendi i limiti.

Ecco una checklist rapida per utilizzare gli LLM in modo responsabile, efficace e sicuro.

Ambito	Che cosa fare	Che cosa evitare
Conoscere lo strumento	Comprendi il modello che usi: nome, dimensioni e scopo.	Usarlo alla cieca senza sapere come funziona.
Formulare prompt chiari	Sii specifico, fornisci contesto e definisci ruoli, per esempio «Sei un insegnante...».	Scrivere prompt vaghi o ambigui.
Pensare in modo critico	Ricontrolla sempre i fatti, soprattutto negli ambiti sensibili.	Fidarti dell’output senza verificarlo.
Prestare attenzione ai bias	Metti alla prova il modello in scenari diversi e da prospettive differenti.	Presumere che il modello sia neutrale o equo per impostazione predefinita.
Proteggere la privacy	Evita di incollare dati personali o riservati nei modelli pubblici.	Presumere che il sistema sia privato o sicuro per impostazione predefinita.
Usare responsabilmente	Considera l’impatto dell’output e le persone coinvolte.	Usare gli LLM per automatizzare compiti senza supervisione umana.
Affinare quando serve	Rivedi prompt e istruzioni in modo iterativo per migliorare i risultati.	Aspettarti risposte perfette al primo tentativo.

Conclusione

Gli LLM stanno cambiando il modo in cui interagiamo con la tecnologia. Possono scrivere, tradurre, riassumere, spiegare e persino sorprenderci con la loro creatività.

Il loro funzionamento si basa su rappresentazioni apprese e generazione di token. Possono svolgere compiti complessi, ma una risposta fluida non dimostra esperienza soggettiva, emozioni o affidabilità. La scelta d’uso deve dipendere da risultati controllati e limiti riconosciuti.

Eppure, quando vengono usati con attenzione, possono essere partner potenti:

- Ci aiutano a risparmiare tempo.
- Rendono più accessibili le idee complesse.
- Sostengono apprendimento, creatività e comunicazione.

La chiave è ricordare che il controllo resta a noi. Gli LLM sono qui per amplificare il nostro lavoro, far nascere idee nuove e rimuovere alcuni ostacoli nelle attività quotidiane, anziché sostituire le persone.

Utilizzati con cura, curiosità e responsabilità, questi modelli possono aiutarci a fare di più e a pensare meglio.

Che tu sia un insegnante, un dirigente d’azienda, uno studente o semplicemente una persona curiosa, ora possiedi le conoscenze di base per confrontarti con questa tecnologia con maggiore sicurezza.

Glossario

Termine	Definizione
IA — Intelligenza artificiale, in inglese AI	Tecnologia che imita le capacità umane di prendere decisioni, risolvere problemi o percepire.
Modello linguistico	Sistema addestrato a prevedere e generare il linguaggio umano sulla base di schemi presenti nel testo.
LLM — Large Language Model	Tipo di IA addestrata su quantità enormi di testo per generare e comprendere il linguaggio.
Prompt	Istruzione o domanda in linguaggio naturale data al modello per guidarne l’output.
Prompt engineering — Ingegneria dei prompt	Pratica di progettare prompt chiari ed efficaci per ottenere risultati migliori dagli LLM.
Token	Porzione di testo, una parola o parte di una parola, che il modello utilizza nella previsione.
Inferenza	Processo con cui un modello addestrato genera una risposta.
Fine-tuning	Ulteriore addestramento di un LLM su dati specialistici per adattarlo a compiti o settori specifici.
Bias	Presupposti o stereotipi involontari riflessi nel modello a causa degli schemi presenti nei dati di addestramento.
Allucinazione	Situazione in cui un LLM genera informazioni false ma convincenti.
Capacità emergente	Prestazione rilevata nei modelli più grandi e non rilevata nei più piccoli in una specifica valutazione; il risultato dipende anche dalla metrica.
Modello multimodale	Modello che può elaborare e generare non soltanto testo, ma anche immagini, audio o altri tipi di dati.
Modello basato sul cloud	LLM utilizzato attraverso Internet, in esecuzione su server remoti.
Modello locale	Modello eseguito nell’ambiente controllato dall’utente; le dimensioni possibili dipendono dall’hardware disponibile.
RAG — Retrieval-Augmented Generation, generazione aumentata dal recupero di informazioni	Tecnica che permette agli LLM di recuperare informazioni da fonti esterne durante la generazione.
Quantizzazione	Metodo per ridurre le dimensioni e l’uso di risorse di un modello, spesso per eseguirlo localmente.

Vuoi saperne di più — Risorse da esplorare

Lecture fondamentali

Attention Is All You Need — L'attenzione è tutto ciò che serve

Vaswani e colleghi, 2017. Il lavoro introduce il transformer e ne valuta le prestazioni nella traduzione. Questa architettura è alla base di molti LLM moderni; il lavoro originale non dimostra le capacità di ogni modello successivo.

<https://arxiv.org/abs/1706.03762>

Man is to Computer Programmer as Woman is to Homemaker — L'uomo sta al programmatore informatico come la donna sta alla casalinga

Bolukbasi e colleghi, 2016. Uno studio sui bias di genere nei word embedding, le rappresentazioni numeriche delle parole, che propone anche un metodo per ridurre alcune manifestazioni. Non è una prova che un intervento elimini ogni forma di bias.

<https://arxiv.org/abs/1607.06520>

On the Dangers of Stochastic Parrots — Sui pericoli dei pappagalli stocastici

Bender e colleghi, 2021. Un'analisi critica dei rischi, dei bias e dell'etica nei modelli linguistici di grandi dimensioni.

<https://dl.acm.org/doi/10.1145/3442188.3445922>

Unione europea — Regolamento sull'intelligenza artificiale

Il regolamento (UE) 2024/1689 è disponibile su EUR-Lex, il portale ufficiale del diritto dell'Unione. Per un impiego concreto consulta il testo vigente, l'ambito applicabile e le scadenze; questa guida non sostituisce una valutazione giuridica. Il sito divulgativo AI Act dell'originale non è il sito della Commissione europea.

<https://eur-lex.europa.eu/eli/reg/2024/1689/oj>

Guide didattiche

OpenAI — Guida all'ingegneria dei prompt

Esempi pratici e suggerimenti per costruire prompt migliori.

<https://platform.openai.com/docs/guides/prompt-engineering>

DeepLearning.AI — ChatGPT Prompt Engineering for Developers — Ingegneria dei prompt per sviluppatori con ChatGPT

Corso breve con esempi per sviluppatori. Verifica sul sito modalità di accesso e prerequisiti, che possono cambiare.

<https://www.deeplearning.ai/short-courses/chatgpt-prompt-engineering-for-developers/>

OCSE – L'IA e il futuro delle competenze

Rapporti sull'impatto dell'IA nell'istruzione e nel lavoro.

<https://www.oecd.org/en/topics/artificial-intelligence-and-education-and-skills.html>

Riferimenti della revisione

Questi riferimenti sostengono le correzioni e le distinzioni aggiunte. Gli esempi con ricevute, matite, tariffe e documenti aziendali sono esercizi originali con dati fittizi. Le fonti non certificano quei dati né le prestazioni di un prodotto specifico.

Emergenza e valutazione

Wei e colleghi, Emergent Abilities of Large Language Models, 2022. Per il rapporto fra scala e prestazioni osservate.

<https://arxiv.org/abs/2206.07682>

Schaeffer, Miranda e Koyejo, Are Emergent Abilities of Large Language Models a Mirage?, 2023. Per il ruolo della metrica nei salti apparenti di prestazione.

<https://arxiv.org/abs/2304.15004>

Spiegazioni e loro limiti

Turpin e colleghi, Language Models Don't Always Say What They Think: Unfaithful Explanations in Chain-of-Thought Prompting, NeurIPS 2023. Per la distinzione fra una spiegazione plausibile e un resoconto fedele delle influenze sul risultato.

https://proceedings.neurips.cc/paper_files/paper/2023/hash/ed3fea9033a80fea1376299fa7863f4a-Abstract.html

Recupero di informazioni

Lewis e colleghi, Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, 2020. Il lavoro combina un modello generativo con il recupero di documenti e valuta compiti che richiedono conoscenza. Non garantisce che ogni fonte recuperata sia aggiornata o corretta.

<https://arxiv.org/abs/2005.11401>

Memoria e trattamento dei dati

OpenAI, documentazione sulle memorie di ChatGPT. Descrive funzioni di un prodotto: non va generalizzata a ogni LLM.

<https://learn.chatgpt.com/docs/customization/memories>

OpenAI, Data controls in the OpenAI platform. Distingue uso per addestramento e conservazione dei dati nelle API. Le condizioni possono cambiare; ricontrollale per il servizio e il contratto applicabili.

<https://developers.openai.com/api/docs/guides/your-data>

Come leggere i casi di cronaca

Le fonti dell'ordinanza Mata, della sentenza colombiana, del rapporto Thomson Reuters Institute e dei resoconti Amazon, Eliza e Hoshi sono riportate accanto ai casi. Le cronache hanno un valore diverso dagli atti giudiziari e dagli studi sperimentali. Per alcuni articoli l'accesso integrale è limitato: la revisione lo segnala, evita attribuzioni non sostenute e conserva la traduzione originale per il confronto. Consultazione dei riferimenti della revisione: 5 ottobre 2026.