Generative AI per Tutti: Appunti del Corso

Il corso "Generative AI for Everyone" si propone di fornire una comprensione accurata e non tecnica dell'intelligenza artificiale generativa, un campo che ha catturato l'attenzione di individui, aziende e governi in tutto il mondo. Dal rilascio di ChatGPT nel novembre 2022, questa tecnologia dirompente ha iniziato a cambiare il modo in cui molte persone apprendono e lavorano, promettendo guadagni di produttività significativi e contributi sostanziali alla crescita economica globale, pur sollevando preoccupazioni riguardo alla perdita di posti di lavoro e ad altri potenziali rischi.

Il corso non richiede alcuna preparazione tecnica o conoscenza preliminare dell'intelligenza artificiale ed è stato progettato per essere utile a chiunque operi in ambito aziendale, scientifico, ingegneristico, umanistico, artistico o in altri settori. L'obiettivo principale è quello di chiarire cosa sia realmente l'intelligenza artificiale generativa, cosa possa e non possa fare, e come utilizzarla efficacemente nel proprio lavoro o nella propria attività.

Secondo le stime di McKinsey, l'intelligenza artificiale generativa potrebbe aggiungere tra i 2,6 e i 4,4 trilioni di dollari all'anno all'economia mondiale. Goldman Sachs prevede che possa aumentare il PIL globale del 7% nel prossimo decennio. Uno studio condotto da OpenAI e dall'Università della Pennsylvania stima che questa tecnologia potrebbe influenzare il 10% del lavoro o delle attività svolte quotidianamente da oltre l'80% dei lavoratori negli Stati Uniti, mentre il 20% dei lavoratori potrebbe vedere più della metà delle proprie mansioni impattate dall'intelligenza artificiale generativa.

L'intelligenza artificiale generativa si riferisce a sistemi di intelligenza artificiale capaci di produrre contenuti di alta qualità, specificamente testo, immagini e audio. Tra i sistemi più noti troviamo ChatGPT di OpenAI, che può seguire istruzioni per svolgere compiti complessi come scrivere didascalie per post sui social media o generare vari output creativi. Oltre alle applicazioni rivolte ai consumatori, come Bard di Google e Bing Chat di Microsoft, esiste un'altra

applicazione dell'intelligenza artificiale generativa che potrebbe rivelarsi ancora più influente nel lungo termine: il suo utilizzo come strumento per sviluppatori.

L'intelligenza artificiale è già pervasiva nelle nostre vite quotidiane. Ogni volta che effettuiamo una ricerca su Google o Bing, utilizziamo la nostra carta di credito per un acquisto, o riceviamo raccomandazioni su Amazon o Netflix, stiamo interagendo con sistemi di intelligenza artificiale. Tuttavia, molti di questi sistemi sono stati complessi e costosi da costruire. L'intelligenza artificiale generativa sta rendendo molto più semplice la creazione di numerose applicazioni di intelligenza artificiale, permettendo alle aziende di sviluppare applicazioni preziose a costi significativamente inferiori rispetto al passato.

Tra i tre tipi di contenuti che l'intelligenza artificiale generativa può produrre, ovvero testo, immagini e audio, l'impatto maggiore si è avuto finora sulla generazione di testo. Tuttavia, la tecnologia può anche generare immagini straordinarie a partire da istruzioni specifiche, creare cloni vocali realistici e persino produrre video che combinano audio e immagini in movimento.

Il corso si articola in tre settimane. Durante la prima settimana, viene esplorato il funzionamento della tecnologia di intelligenza artificiale generativa, con particolare attenzione a cosa può e non può fare, presentando una varietà di casi d'uso comuni che possono stimolare la creatività nell'applicazione di questa tecnologia. La seconda settimana si concentra sui progetti di intelligenza artificiale generativa, spiegando come identificare applicazioni utili e quali siano le migliori pratiche per svilupparle, approfondendo le varie opzioni tecnologiche disponibili. La terza settimana allarga la prospettiva oltre i singoli progetti per esaminare come l'intelligenza artificiale generativa influenzerà le imprese e la società nel suo complesso, presentando le migliori pratiche per i team aziendali e analizzando i rischi associati all'intelligenza artificiale e come garantire un uso responsabile di questa tecnologia.

Per comprendere come funziona la generazione di testo, è importante collocare l'intelligenza artificiale generativa all'interno del più ampio panorama dell'intelligenza artificiale. L'intelligenza artificiale può essere concepita come un insieme di strumenti, tra cui l'apprendimento supervisionato rappresenta uno dei più importanti. L'apprendimento supervisionato eccelle nell'etichettare le cose, ovvero nel prendere un input A e generare un corrispondente output B. Questo approccio ha reso i computer molto efficaci in compiti come il filtraggio dello

spam, la pubblicità online, la guida autonoma, la diagnosi medica, l'ispezione di difetti di produzione, il riconoscimento vocale e l'analisi del sentiment delle recensioni.

Il decennio tra il 2010 e il 2020 è stato caratterizzato dall'apprendimento supervisionato su larga scala, che ha posto le basi per l'intelligenza artificiale generativa moderna. Durante questo periodo, i ricercatori hanno scoperto che alimentando modelli di intelligenza artificiale molto grandi con enormi quantità di dati, le prestazioni continuavano a migliorare in modo significativo. Questo principio ha guidato lo sviluppo di modelli linguistici di grandi dimensioni, comunemente abbreviati come LLM.

I modelli linguistici di grandi dimensioni generano testo utilizzando l'apprendimento supervisionato per prevedere ripetutamente quale sarà la prossima parola in una sequenza. Quando si addestra un sistema di intelligenza artificiale molto grande su centinaia di miliardi di parole, in alcuni casi anche più di un trilione, si ottiene un modello linguistico di grandi dimensioni come ChatGPT, capace di generare parole aggiuntive in risposta a un prompt fornito. Una singola frase letta su Internet viene trasformata in molteplici punti dati per l'apprendimento, dove il modello impara a prevedere la parola successiva dato un contesto di poche parole precedenti.

Esistono diverse interfacce web attraverso le quali è possibile accedere ai modelli linguistici di grandi dimensioni. ChatGPT è la più nota, ma anche Bard di Google, Microsoft Bing e molti altri funzionano efficacemente. Questi sistemi offrono un nuovo modo di trovare informazioni, sebbene sia importante notare che i modelli linguistici di grandi dimensioni possono talvolta inventare fatti, un fenomeno chiamato allucinazione. Quando si fa affidamento su una risposta accurata, è utile verificare la risposta con una fonte autorevole prima di utilizzarla.

I modelli linguistici di grandi dimensioni possono anche fungere da partner di riflessione per aiutare a elaborare idee. Molte persone li utilizzano per affinare la propria scrittura, riformulare testi per maggiore chiarezza o creare contenuti creativi. Tuttavia, è importante comprendere quando utilizzare una ricerca web tradizionale rispetto a un modello linguistico di grandi dimensioni. Per questioni mediche o ricette consolidate, le fonti web tradizionali provenienti da siti affidabili potrebbero essere più appropriate. Al contrario, per compiti più esoterici o creativi

che richiedono combinazioni insolite di concetti, i modelli linguistici di grandi dimensioni possono essere particolarmente utili come partner di pensiero.

L'intelligenza artificiale generativa è una tecnologia a scopo generale, simile all'elettricità o a Internet, il che significa che è utile per molti compiti diversi piuttosto che per un singolo scopo specifico. Questa caratteristica la rende quasi difficile da descrivere in modo esaustivo. Un framework utile per organizzare i compiti che i modelli linguistici di grandi dimensioni possono svolgere include tre categorie principali: scrittura, lettura e conversazione.

Per quanto riguarda la scrittura, i modelli linguistici di grandi dimensioni sono particolarmente efficaci nel generare testo a partire da prompt relativamente brevi. Possono essere utilizzati come compagni di brainstorming per generare nomi creativi per prodotti, scrivere comunicati stampa, tradurre testi in diverse lingue o creare contenuti creativi. Quando si fornisce maggiore contesto o informazioni di background al modello, questo produce output più specifici e di qualità superiore. Per esempio, un prompt generico per scrivere un comunicato stampa produrrà un risultato generico, mentre fornire dettagli specifici sull'azienda e sulla persona coinvolta genererà un testo molto più pertinente e professionale.

La traduzione è un'altra applicazione importante della scrittura. I modelli linguistici di grandi dimensioni sono competitivi e talvolta superiori ai motori di traduzione dedicati, specialmente per le lingue con una grande quantità di testo disponibile su Internet. Tuttavia, tendono a funzionare meno bene per le lingue con meno risorse online. Un approccio interessante utilizzato nella comunità dell'intelligenza artificiale per testare le traduzioni è quello di tradurre in inglese piratesco, che permette di verificare rapidamente se il modello sta producendo traduzioni ragionevoli anche per chi non parla la lingua di destinazione.

Le attività di lettura rappresentano la seconda categoria principale. In questi compiti, si fornisce al modello un testo relativamente lungo e si chiede di generare un output più breve. Un esempio comune è la correzione di bozze, dove i modelli linguistici di grandi dimensioni possono identificare errori di ortografia, grammatica e frasi poco chiare che potrebbero sfuggire anche a una lettura attenta. La sintesi di articoli lunghi è un'altra applicazione preziosa, permettendo di comprendere rapidamente il contenuto di documenti estesi senza doverli leggere interamente.

Le applicazioni di lettura basate su software stanno diventando sempre più comuni nelle aziende. Per esempio, nei call center del servizio clienti, le conversazioni telefoniche possono essere trascritte e poi sintetizzate automaticamente dai modelli linguistici di grandi dimensioni, permettendo ai manager di rivedere rapidamente molte interazioni con i clienti. L'analisi delle email dei clienti è un'altra applicazione importante, dove i modelli possono classificare le email come reclami o meno e indirizzarle al dipartimento appropriato. Per ottenere risultati accurati in questi casi, è essenziale fornire al modello un contesto sufficiente, come l'elenco dei dipartimenti effettivamente esistenti nell'organizzazione.

Il monitoraggio della reputazione è un'altra applicazione pratica delle attività di lettura. Le aziende possono utilizzare i modelli linguistici di grandi dimensioni per analizzare automaticamente le recensioni online e classificarle come positive o negative, creando dashboard che tracciano l'andamento del sentiment dei clienti nel tempo. Questo permette di identificare rapidamente tendenze preoccupanti che potrebbero richiedere attenzione.

Le attività di conversazione costituiscono la terza categoria principale. Oltre ai chatbot generici come ChatGPT, Bard e Bing Chat, molte aziende stanno esplorando la possibilità di costruire chatbot specializzati per scopi specifici. Questi possono includere chatbot per il servizio clienti che prendono ordini, chatbot per la pianificazione di viaggi, chatbot di consulenza per varie questioni o chatbot di supporto IT che gestiscono richieste comuni come il reset delle password.

Esistono diversi approcci per implementare i chatbot nei centri di assistenza clienti. All'estremità dello spettro si trovano i centri con solo agenti umani, mentre all'altra estremità ci sono sistemi completamente automatizzati. Tra questi estremi esistono diverse configurazioni comuni. Una soluzione intermedia prevede che i bot supportino gli umani generando messaggi suggeriti che gli agenti possono rivedere, modificare e approvare prima dell'invio. Questo approccio, chiamato human-in-the-loop, mitiga il rischio che il chatbot dica qualcosa di inappropriato. Un'altra configurazione prevede che i bot gestiscano i messaggi semplici e inoltrino ai dipendenti umani solo i casi più complessi o delicati.

Molte aziende adottano un approccio graduale nell'implementazione dei chatbot. Inizialmente, possono sviluppare un chatbot rivolto solo all'interno, permettendo al

proprio team di utilizzarlo e testarlo in un ambiente controllato dove eventuali errori sarebbero più facilmente perdonati. Successivamente, possono implementare il sistema con supervisione umana, dove gli agenti controllano i messaggi prima che vengano inviati ai clienti. Infine, quando il sistema dimostra di essere sufficientemente affidabile, può essere autorizzato a comunicare direttamente con i clienti. Naturalmente, i dettagli specifici variano a seconda del business, del volume di traffico e dei rischi associati a eventuali errori del bot.

È importante comprendere cosa i modelli linguistici di grandi dimensioni possono e non possono fare. Un framework mentale utile consiste nel chiedersi se un neolaureato, seguendo solo le istruzioni contenute nel prompt, potrebbe completare il compito richiesto. Questo neolaureato ipotetico ha una vasta conoscenza generale acquisita da Internet, ma non ha accesso a motori di ricerca web e non sa nulla di specifico sulla vostra azienda o attività. Inoltre, ogni volta che si utilizza un modello linguistico di grandi dimensioni, è come se si ottenesse un neolaureato diverso, poiché il modello non ricorda le conversazioni precedenti.

Questo modello mentale, sebbene imperfetto, fornisce un punto di partenza utile per comprendere le capacità e i limiti dei modelli linguistici di grandi dimensioni. Esistono cose che i neolaureati possono fare che i modelli non possono fare e viceversa, ma il parallelo aiuta a formare aspettative realistiche su ciò che si può ottenere attraverso il prompting.

I modelli linguistici di grandi dimensioni presentano diverse limitazioni specifiche che è importante conoscere. Il primo limite riguarda il cutoff della conoscenza. La conoscenza del mondo di un modello è congelata al momento del suo addestramento. Un modello addestrato su dati di Internet raccolti fino a gennaio 2022 non avrà informazioni su eventi successivi. Questo significa che non potrà rispondere a domande su film di successo del 2022 o eventi scientifici avvenuti dopo quella data.

Le allucinazioni rappresentano una seconda limitazione significativa. I modelli linguistici di grandi dimensioni talvolta inventano informazioni con tono molto sicuro e autorevole. Possono creare citazioni false attribuite a personaggi storici o elencare casi giudiziari inesistenti. Questo fenomeno può avere conseguenze gravi, come dimostrato dal caso di un avvocato che ha presentato in tribunale documenti legali contenenti casi giudiziari inventati da ChatGPT, venendo poi sanzionato per aver sottoposto materiale fabbricato.

I modelli linguistici di grandi dimensioni hanno anche limitazioni tecniche relative alla lunghezza dell'input e dell'output. Molti modelli possono accettare prompt fino a poche migliaia di parole, limitando la quantità totale di contesto che si può fornire. Per documenti più lunghi, potrebbe essere necessario suddividere il contenuto e processarlo in parti separate, oppure cercare modelli con limiti di input più elevati che possono gestire decine di migliaia di parole.

Una limitazione importante riguarda i dati strutturati. I modelli linguistici di grandi dimensioni non funzionano bene con dati tabulari, come quelli che si potrebbero memorizzare in un foglio di calcolo Excel o Google Sheets. Per compiti che coinvolgono la previsione di valori basati su dati tabulari, come stimare il prezzo di una casa in base alle sue caratteristiche o prevedere il comportamento di acquisto degli utenti, l'apprendimento supervisionato tradizionale rappresenta un approccio più appropriato rispetto all'utilizzo di modelli linguistici di grandi dimensioni.

I modelli linguistici di grandi dimensioni tendono invece a funzionare meglio con dati non strutturati, che includono testo, immagini, audio e video. Sebbene l'intelligenza artificiale generativa si applichi a tutti questi tipi di dati, l'impatto maggiore si è avuto finora con i dati testuali.

Un'altra preoccupazione importante riguarda i bias e il linguaggio dannoso. I modelli linguistici di grandi dimensioni sono stati addestrati su testi provenienti da Internet, che possono riflettere pregiudizi esistenti nella società. Ad esempio, potrebbero associare automaticamente determinate professioni a specifici generi, assumendo che i chirurghi siano maschi e le infermiere siano femmine. Quando si utilizzano questi modelli in applicazioni dove tali bias potrebbero causare danni, è necessario prestare particolare attenzione al modo in cui si formulano i prompt e si applicano i modelli.

Alcuni modelli possono occasionalmente produrre contenuti tossici o dannosi, incluse istruzioni su come compiere atti indesiderabili o illegali. Fortunatamente, tutti i principali fornitori di modelli linguistici di grandi dimensioni stanno lavorando intensamente sulla sicurezza di questi modelli, che sono diventati molto più sicuri nel tempo. Le interfacce web dei principali fornitori di modelli hanno implementato misure che rendono sempre più difficile ottenere output dannosi.

Per ottenere i migliori risultati quando si utilizzano i modelli linguistici di grandi dimensioni, esistono alcune pratiche consigliate per la formulazione dei prompt. Il primo suggerimento è essere dettagliati e specifici. Utilizzando l'analogia del

neolaureato, è importante assicurarsi che il modello abbia contesto e informazioni di background sufficienti per completare il compito. Fornire dettagli rilevanti e descrivere chiaramente cosa si desidera ottenere aumenta significativamente la probabilità di ricevere un risultato soddisfacente.

Il secondo suggerimento consiste nel guidare il modello attraverso il processo di ragionamento. Se si ha in mente un processo specifico che il modello dovrebbe seguire per arrivare alla risposta desiderata, fornire istruzioni chiare passo dopo passo può essere molto efficace. Ad esempio, quando si chiede di generare nomi per giocattoli per gatti con rime ed emoji, specificare i passaggi esatti che il modello dovrebbe seguire porta a risultati migliori.

Il terzo suggerimento riguarda la sperimentazione e l'iterazione. Non esiste un prompt perfetto universale per tutti. Piuttosto che cercare il prompt ideale dall'inizio, è più produttivo iniziare con qualcosa di ragionevolmente chiaro e specifico, anche se breve, e poi raffinare il prompt in base ai risultati ottenuti. Il processo di prompting è altamente iterativo: si inizia con un'idea di cosa si vuole che il modello faccia, si formula un prompt, si osserva la risposta e, se non è soddisfacente, si modifica il prompt basandosi sul motivo per cui il risultato non corrisponde alle aspettative.

È importante non preoccuparsi eccessivamente della formulazione iniziale del prompt. È meglio provare rapidamente qualcosa e, se non fornisce i risultati desiderati, migliorarlo nel tempo. Non si danneggia Internet con un prompt formulato in modo leggermente errato. Tuttavia, esistono due importanti avvertenze: prima di incollare informazioni altamente riservate nell'interfaccia web di un modello linguistico, è necessario comprendere come il fornitore utilizza e protegge tali informazioni. In secondo luogo, come dimostrato dal caso dell'avvocato che ha sottoposto documenti con fatti inventati, prima di fare affidamento sull'output di un modello, può essere opportuno verificare e decidere autonomamente se si può confidare in quella informazione e agire di conseguenza.

Per quanto riguarda la generazione di immagini, questa rappresenta un'altra importante applicazione dell'intelligenza artificiale generativa. Alcuni modelli possono generare sia testo che immagini e sono chiamati modelli multimodali perché possono operare in più modalità. La generazione di immagini viene principalmente realizzata attraverso un metodo chiamato modello di diffusione.

I modelli di diffusione hanno appreso da enormi quantità di immagini trovate su Internet. Al cuore di un modello di diffusione si trova l'apprendimento supervisionato. Il processo funziona nel modo seguente: prendendo un'immagine, ad esempio di una mela, il modello aggiunge gradualmente sempre più rumore all'immagine fino a ottenere un'immagine di puro rumore casuale. Questo processo crea una serie di immagini con livelli crescenti di rumore.

Il modello di diffusione utilizza quindi queste immagini come dati per apprendere, attraverso l'apprendimento supervisionato, a prendere un'immagine rumorosa e produrre un'immagine leggermente meno rumorosa. Dopo l'addestramento su centinaia di milioni di immagini attraverso questo processo, quando si vuole generare una nuova immagine, si inizia con un'immagine di puro rumore casuale. Questa viene alimentata all'algoritmo di apprendimento supervisionato addestrato, che rimuove progressivamente il rumore attraverso circa cento passaggi successivi, fino a ottenere un'immagine chiara.

Per controllare il tipo di immagine generata attraverso un prompt testuale, il modello viene modificato durante l'addestramento. Invece di fornire solo l'immagine rumorosa come input, si fornisce sia l'immagine rumorosa che una descrizione testuale dell'immagine originale. Durante la generazione, partendo da puro rumore e fornendo un prompt come "banana verde", l'algoritmo rimuove progressivamente il rumore mantenendo la coerenza con il prompt fornito, fino a produrre un'immagine che corrisponde alla descrizione richiesta.

Questo processo magico di generazione di immagini belle e complesse si basa, ancora una volta, sull'apprendimento supervisionato come tecnologia fondamentale. La combinazione di questi metodi sta aprendo nuove possibilità creative e applicazioni pratiche in numerosi settori, dalla progettazione grafica alla prototipazione di prodotti, dall'arte digitale alla visualizzazione di concetti astratti.