Intelligenza artificiale: il caso ChatGPT

ChatGPT tra sperimentazione, dubbi e burocratici divieti

Cresce l'interesse, anche nel mondo delle biblioteche, per l'intelligenza artificiale (IA). In particolare, nell'ultimo periodo, l'attenzione si è focalizzata su ChatGPT, chatbot che ha suscitato contrastanti giudizi e reazioni sulla sua capacità di garantire informazioni corrette. Anche i giornali hanno cominciato a occuparsene ed è proprio dalla cronaca che apprendiamo mentre stiamo andando in stampa che il servizio è stato "oscurato" dal Garante della privacy.

Per i bibliotecari, che considerano l'informazione una risorsa strategica su cui esercitare la propria professionalità, è importante affrontare l'argomento.

Per questo la nostra rivista intende avviare da questo numero un percorso che ci accompagnerà attraverso approfondimenti, analisi ed esperienze a maturare un quadro più definito ma pur sempre aperto al confronto.

Uno degli scopi principali con cui OpenAI nel dicembre 2022 ha reso pubblicamente accessibile ChatGPT è stata l'esigenza di coinvolgere il maggior numero di persone nel prendere contatto con questa nuova tecnologia, poterla sperimentate, poterne dibattere.

Il risultato è stato ampiamente raggiunto, da allora infatti decine di milioni di persone in tutto il mondo hanno potuto interagire con ChatGPT e ne è scaturito un esteso dibattito pubblico ricco di opinioni, entusiasmi, critiche, necessità di approfondimento con implicazioni sociali, commerciali e industriali.

La tecnologia nel frattempo è migliorata anche in risposta agli esperimenti pubblicamente condotti e ogni giorno si affacciano decine di nuovi strumenti basati sulle nuove capacità delle tecnologie alla base di ChatGPT.

Il dibattito si è esteso ed evoluto portando a diverse posizioni come, ad esempio, a fine marzo 2023, la lettera aperta del Future of Life Institute sottoscritta da migliaia di esponenti della ricerca scientifica e tecnologica dove viene richiesto di sospendere per sei mesi lo sviluppo di sistemi di AI più potenti dell'attuale GPT-4 per consentire alla società di comprendere meglio i rischi e

le potenzialità di queste nuove tecnologie e i metodi di controllo e garanzia degli impatti benefici.
Di altro tono e caso più eccezionale, in Italia, il dibattito è giunto a un brusco arresto per cui il servizio ChatGPT, in questo momento, non è disponibile per gli utenti che si collegano dall'Italia. OpenAI per rispondere a una istruttoria mossa del Garante della privacy ha dovuto sospendere il servizio per noi italiani.
Riteniamo che i dubbi sollevati dal Garante non vadano sottovalutati e debbano indurre a valutare meglio anche i temi legati alla privacy e al trattamento dei dati personali. Tuttavia ci sembra che la modalità così brusca, la richiesta eccessiva e probabilmente anche la non piena consapevolezza dello strumento, dello scopo e del suo funzionamento che è stata mossa dal Garante non rappresenti il modo migliore per favorire questo tipo di riflessione. L'accesso a queste tecnologie è di fondamentale importanza per poterle conoscere, studiare, comprendere e quindi contribuire alla loro evoluzione.
Vedremo negli articoli che seguono una rassegna di esperienze e riflessioni, a partire dal primo di questi contributi che si propone di delineare un quadro d'insieme circa la rilevanza e l'interesse che hanno i *large language model* e i recenti progressi dell'intelligenza artificiale in cui si situa ChatGPT e di cui rappresenta una modalità di contatto e di interazione.

È iniziata l'era dell'intelligenza artificiale

ChatGPT e i primi segni della nuova rivoluzione

LORENZO VERNA
v.lorenzo@gmail.com

DOI: 10.3302/0392-8586-202303-004-1

Il titolo ipotizzato per questo breve contributo era *Intelligenza Artificiale: l'era dell'adozione*. In fase di stesura inoltrata ho ricevuto la notifica della pubblicazione di un nuovo articolo di Bill Gates sul tema, *The age of AI has begun*, con il sottotitolo *Artificial intelligence is as revolutionary as mobile phones and the Internet*.[1]
Da professionista che si occupa in varie forme di Intelligenza Artificiale (AI) da più di vent'anni percepisco netta la sensazione che il periodo che stiamo vivendo sia eccezionale, che assomigli a un rinnovamento epocale; quindi, sostenuto anche dall'autorevolezza di Bill Gates, ho ritenuto che il titolo del suo contributo esprima meglio quello che più timidamente voleva indicare il titolo che avevo immaginato. Nel suo articolo Gates si interroga e propone possibili percorsi su come nel prossimo futuro l'evoluzione dell'Intelligenza Artificiale (IA) potrà contribuire a risolvere i problemi più profondi che ancora oggi affliggono l'umanità e di cui lui si occupa a tempo pieno con la Bill & Melinda Gates Foundation. Ritengo interessante riportare alcuni passaggi della sua introduzione, dove espone le ragioni che lo hanno indotto a ritenere questa che stiamo vivendo una effettiva rivoluzione tecnologica.
Nella mia vita ho assistito a due dimostrazioni di tecnologia che mi hanno colpito per essere rivoluzionarie. La prima volta fu nel 1980 quando mi presentarono un'interfaccia utente grafica, il precursore di ogni moderno sistema operativo, incluso Windows [...] La seconda grande sorpresa l'ho avuta proprio l'anno scorso. Dal 2016 mi incontro regolarmente con il team di OpenAI e mi hanno sempre impressionato per la rapidità dei loro progressi. A metà 2022 ero così entusiasta del loro lavoro che proposi una sfida: addestrare un'intelligenza artificiale per superare l'esame Advanced Placement in Biology. Dovevano renderla in grado di rispondere a domande per le quali non è stata specificatamente addestrata. (Ho scelto AP in Biologia perché il test è molto più che una semplice restituzione di nozioni scientifiche, richiede invece un pensiero critico sulla biologia). Se riuscirete a farlo, dissi, allora avrete ottenuto un si-

gnificativo passo avanti. Pensai che il problema proposto li avrebbe tenuti occupati due o tre anni. Finirono in pochi mesi. A settembre, quando li incontrai di nuovo, osservai con ammirazione mentre loro chiedevano a GPT, il loro modello di IA, 60 domande a risposta multipla dell'esame AP Bio, e GPT rispose correttamente a 59. Quindi scrisse ottime risposte alle sei domande aperte dell'esame. Era presente un esperto esterno per valutare l'esito del test, e GPT ha ottenuto 5 – il punteggio più alto possibile, l'equivalente di prendere A o A+ a un corso di biologia di livello universitario. Una volta superato l'esame, gli abbiamo chiesto una domanda non scientifica: "Cosa diresti al padre di un figlio malato?". GPT ha scritto una risposta premurosa probabilmente migliore di quella che tutti noi nella stanza avremmo dato. L'intera esperienza è stata sbalorditiva.
Sapevo di aver appena assistito al più importante progresso tecnologico dall'avvento delle Interfacce Utente Grafiche (Graphical User Interfaces - GUI).[2]

1. Mesi di fervore

Nell'ultimo anno i recenti progressi nell'AI si susseguono con un ritmo sempre crescente, per cui è diventato difficile anche solo orientarsi tra le innumerevoli novità che ormai quotidianamente ridisegnano il panorama del settore, sia dal punto di vista scientifico che delle applicazioni, dei prodotti commerciali, degli assetti societari e degli impatti sul mercato e sull'economia.
Ad esempio, anche semplicemente tratteggiare una mappa, elaborare il materiale per un seminario o una pubblicazione risulta molto complicato in quanto nel tempo stesso dell'ideazione, redazione e pubblicazione i fatti che si raccontano sono invalidati dalle novità che inesorabilmente vengono annunciate introducendo elementi di novità che modificano anche sostanzialmente il contesto e lo scenario esaminato. In questa ottica possiamo sostenere che una cosa è certa: il periodo storico che stiamo vivendo è eccezionale, se non altro per la tumultuosità crescente con cui i nuovi elementi vengono prodotti, annunciati e resi disponibili dalla ricerca e dall'industria di settore.
Chi volesse avere un'idea più concreta del ritmo e della portata con cui le novità si susseguono si può ad esempio iscrivere a *Ben's Bites* (https://www.bensbites.co), una newsletter gratuita in lingua inglese redatta da Ben Tossell, un professionista indipendente che tenta di raccogliere quotidianamente le notizie più rilevanti.
L'aspetto rivoluzionario che l'attuale tecnologia di AI lascia intravedere, oltre ai risultati straordinari raggiunti come quelli, ad esempio, citati da Bill Gates, è anche determinato dal progressivo incremento della possibilità di adozione in svariati sistemi e processi. Fino a pochi anni fa l'introduzione di capacità di AI avanzate all'interno di un'applicazione o un sistema richiedeva un ingente investimento di tempo e risorse a cui spesso non corrispondevano risultati adeguati. Oggi assistiamo a una diffusione quasi pervasiva di certe funzionalità e ciò è principalmente dovuto a due fattori. Un primo elemento di accelerazione alla diffusione di funzionalità abilitate dall'AI in svariate applicazioni digitali è data dalla possibilità di integrare in modo relativamente semplice ed economicamente sostenibile servizi e sistemi sviluppati da grandi aziende specializzate che hanno investito ingenti risorse (denaro, risorse computazionali, competenze, dati, collaborazioni con altri istituti di ricerca ecc.) rendendole disponibili ad altri progettisti per la realizzazione di nuovi prodotti specifici. Un secondo fattore molto importante è la crescente possibilità di addestrare modelli per casi d'uso specifici con dati proprietari. Quest'ultima prospettiva è sempre stata percorribile, ma aveva una soglia di ingresso molto alta e pochi potevano utilizzarla. Oggi invece, con la possibilità di condividere modelli già pre-addestrati, la disponibilità di piattaforme di sviluppo specifiche, software, dati e strumenti Open Source, questa pratica si va diffondendo ponendo in essere un processo di adozione su larga scala e di democratizzazione dell'AI, e per questo si tratta di un processo molto importante per riequilibrare l'attuale asimmetria rispetto ai giganti della tecnologia.
Nei recenti anni, mesi e settimane il campo dell'AI ha assistito a progressi significativi in particolare nello sviluppo della *generative AI* e dei Large Language Models (LLMs) come il Generative Pre-trained Transformer (GPT) e le sue varianti, GPT-3[3] e GPT-4.[4] Questi modelli hanno dimostrato capacità sorprendenti nel processamento del linguaggio naturale e nell'interazione con gli utenti, aprendo nuove possibilità per l'automazione e il miglioramento di vari processi in diversi settori.
L'obiettivo di questo contributo è tentare di offrire una panoramica dello scenario in rapida evoluzione in cui si situa GPT e il contributo che esso porta al progresso della disciplina.

2. Intelligenza Artificiale Generativa e LLM: una breve panoramica

L'intelligenza artificiale generativa (*generative AI*) si riferisce alla classe di modelli di AI in grado di produrre contenuti nuovi e originali, come testi, immagini e musica, a partire da un insieme di dati di input. Questi modelli di apprendimento automatico sono in grado di generare output che imitano lo stile e la struttura dei dati di input, aprendo la strada a una vasta gamma di applicazioni pratiche.

2.1 Modelli generativi

I modelli generativi sono algoritmi di apprendimento automatico che cercano di apprendere la distribuzione dei dati di input e di generare nuovi dati che seguono la stessa distribuzione. Alcuni esempi di modelli generativi includono le reti neurali generative avversarie (GAN), i modelli di Boltzmann, i campi casuali di Markov e i *variational autoencoder* (VAE).
Le GAN sono una classe di modelli generativi introdotta nel 2014 da Ian Goodfellow[5] che utilizzano due reti neurali distinte: un generatore e un discriminatore. Il generatore crea nuovi dati, mentre il discriminatore cerca di distinguere tra i dati reali e quelli generati. Le due reti vengono addestrate in modo competitivo, con il generatore che cerca di ingannare il discriminatore e il discriminatore che cerca di migliorare la sua capacità di distinguere tra i dati reali e quelli generati. Questo processo porta a un miglioramento iterativo delle performance di entrambe le reti.

2.2 Applicazioni di generative AI

I modelli generativi trovano applicazione in una vasta gamma di settori, come la generazione di immagini, la sintesi di testo, la creazione di musica, il miglioramento delle immagini, la traduzione tra domini e la generazione di dati per l'addestramento di altri modelli di *machine learning*.

2.2.1 Sintesi di immagini

Nel campo della *generative AI* le tecniche più utilizzate per la generazione di immagini sono:

1. GANs (*Generative Adversarial Networks*): le GANs, come accennato in precedenza, sono architetture composte da due reti neurali, un generatore e un discriminatore, che lavorano in competizione l'uno con l'altro. Il generatore crea immagini false, mentre il discriminatore tenta di distinguere tra immagini false e immagini reali. Durante l'addestramento, il generatore cerca di migliorare la qualità delle immagini create per ingannare il discriminatore, mentre il discriminatore cerca di migliorare la sua capacità di riconoscere le immagini false.
2. VAEs (*Variational autoencoders*): i VAEs sono modelli generativi probabilistici basati sull'*autoencoder*, una rete neurale che impara a comprimere i dati in uno spazio latente e a ricostruirli successivamente. I VAEs aggiungono un vincolo sulla distribuzione dello spazio latente per garantire che le immagini generate siano più variate e realistiche. Durante l'addestramento, i VAEs apprendono a mappare i dati di input in uno spazio latente e a campionare da esso per generare nuove immagini.
3. *Transformer-based models*: l'esempio più noto è quello di DALL-E, un modello di generazione di immagini sviluppato da OpenAI. Questo modello è in grado di creare immagini di alta qualità a partire da descrizioni testuali. A differenza delle GANs e delle VAEs, DALL-E utilizza l'architettura Transformer (un tipo specifico di rete neurale), che è stata inizialmente sviluppata per il trattamento del linguaggio naturale, per modellare le relazioni tra i *token* di input (testo e immagine) e generare nuove immagini.
4. *Denoising diffusion probabilistic models* (DDPM): i DDPM sono modelli generativi che utilizzano un processo di diffusione inversa per generare nuove immagini. Invece di utilizzare un processo avversario come nelle GANs, i DDPM si fondano su un'ottimizzazione basata sulla massima verosimiglianza. Durante l'addestramento, i DDPM apprendono a rimuovere progressivamente il rumore dalle immagini corrotte per ricostruire le immagini originali, e possono generare nuove immagini invertendo questo processo di *denoising*.

Queste tecniche rappresentano diversi approcci alla generazione di immagini e sono state utilizzate con successo in vari ambiti, come:

- la sintesi di immagini a partire da un input testuale, come da esempio in Figura 1: sono disponibili online interi archivi di immagini generate da istruzioni testuali, si veda ad esempio https://www.midjourney.com/showcase/top;
- il completamento di immagini;
- il trasferimento di stile e l'arte generativa: queste tecniche possono essere utilizzate per applicare lo stile di un'immagine a un'altra, creando opere d'arte uniche e personalizzate;
- il miglioramento della risoluzione delle immagini (*super-resolution*): alcune tecniche di generazione di immagini possono essere utilizzate per aumentarne la risoluzione, migliorando la qualità di quelle a bassa risoluzione;
- la generazione di dati per l'addestramento: le tecniche di generazione di immagini possono essere usate per creare grandi set di dati di immagini sintetiche, che possono essere utilizzati per addestrare modelli di apprendimento automatico in scenari in cui i dati reali sono limitati o costosi da ottenere;
- la traduzione di immagini tra diversi domini: alcune tecniche possono essere impiegate per convertire immagini tra diversi domini, ad esempio trasforma-

Figura 1 - Esempio di immagini generate da tre sistemi automatici differenti a fronte dello stesso testo in input: "Un uomo anziano che riposa la testa su un succoso cheeseburger, alta qualità, stile fotografico, ultra realistico, profondità di campo"

re immagini in stile schizzo in immagini a colori o convertire immagini notturne in diurne;
- la generazione di contenuti per videogiochi e realtà virtuale: le tecniche di generazione di immagini possono essere utilizzate per creare automaticamente texture, oggetti e ambienti per videogiochi e applicazioni di realtà virtuale;
- la modellazione e la visualizzazione di dati scientifici: queste tecniche possono essere utilizzate per generare rappresentazioni visive di dati complessi, ad esempio, per creare immagini di molecole o strutture cellulari in base ai dati sperimentali.

Su questi argomenti esiste una vasta letteratura di riferimento, che ne descrivere i complessi caratteri e funzioni.[6]

2.2.2 Creazione di musica

I più recenti progressi nell'AI generativa applicata alla musica sono determinati dallo sviluppo di modelli di AI che possono generare musica di ogni genere data una descrizione testuale. Questi modelli sono in grado di generare musica che è indistinguibile da musica composta da noi umani. Alcuni esempi sono:
- MusicLM, un sistema di AI generativa sviluppato da Google che può creare musica di ogni genere data una descrizione testuale. È in grado di generare brani complessi e in alta fedeltà, e può anche generare variazioni su melodie esistenti;[7]
- Musi-co, una piattaforma per la creazione di musica che permette agli utenti di generare infinite melodie, ritmi e armonie partendo da diversi input, testo, gesti, melodie preesistenti;[8]
- Riffusion, un sistema di AI che compone musica utilizzando un sistema di Stable Diffusion per creare l'immagine dello spettrogramma della melodia che sta generando.[9]

I progressi nella *generative AI* stanno avendo impatti significativi nell'industria musicale rendendo possibile per gli artisti creare nuova musica in modo semplice e con nuove vie per esprimere la propria creatività. Molti compositori stanno già usando queste tecnologie per produrre le loro opere, come il brano *Marry me* dell'artista Robbie Barrat, pubblicato da Columbia Records, o l'album *AI-Generated Music* di Holly Herndon, pubblicato da Warp Records.
Con il progredire di queste tecnologie si intravede la possibilità che aumenterà anche la loro diffusione e il loro impatto nella produzione musicale.

2.2.3 Generazione di testo

La *generative AI* è ampiamente utilizzata anche nella generazione di testo. Modelli come GPT di OpenAI (Brown et al., 2020) sono in grado di produrre testi coerenti e paragonabili a testi scritti da umani in vari stili e formati. Questi modelli possono essere utilizzati per la scrittura creativa, la traduzione automatica, la sintesi di testo e altre applicazioni legate al linguaggio naturale.[10]
I modelli linguistici di grandi dimensioni (*Large Language Models,* LLM) sono un esempio di questo tipo di AI e rappresentano uno dei più grandi progressi nel campo del *Natural Language Processing* (NLP) degli ultimi anni. Gli LLM, come GPT-3, sono potenti modelli di AI generativa che utilizzano enormi quantità di dati

per produrre testi spesso indistinguibili dai contenuti scritti dall'uomo. Questi modelli sono stati addestrati su diverse fonti testuali, consentendo loro di generare informazioni contestualmente rilevanti su una vasta gamma di argomenti.
Vedremo nei prossimi paragrafi come i *Large Language Models* offrono prestazioni tali da renderli strumenti utili in svariate applicazioni, ponendo le basi per la nuova era dell'intelligenza artificiale.

3. Lo sviluppo dei grandi modelli di linguaggio

Negli ultimi anni, i grandi modelli di linguaggio (LLM) sono diventati sempre più importanti nel campo della ricerca sull'intelligenza artificiale (AI), dimostrando la loro capacità di affrontare una vasta gamma di compiti complessi basati sul linguaggio. Questo progresso è stato alimentato da numerosi fattori, tra cui un aumento del numero di parametri del modello, una maggiore quantità di dati di addestramento e una migliore configurazione del training (Brown et al., 2020; Radford et al., 2019; Hernandez et al., 2021; Kaplan et al., 2020). I LLM di ultima generazione, come LaMDA (Thoppilan et al., 2022) e GPT-4 (OpenAI, 2023b), eccellono in diverse applicazioni come la traduzione, la classificazione, la scrittura creativa e la generazione di codice – capacità che in precedenza richiedevano modelli specializzati per compiti specifici sviluppati da ingegneri esperti utilizzando dati specifici del dominio.
Contemporaneamente, i ricercatori hanno migliorato la gestibilità, l'affidabilità e l'utilità di questi modelli utilizzando metodi come il *fine-tuning* e il *reinforcement learning* con feedback umano (Ouyang et al., 2022; Bai et al., 2022). Questi progressi migliorano la capacità dei modelli di comprendere l'intento dell'utente, rendendoli più amichevoli e pratici. Inoltre, recenti studi rivelano il potenziale dei LLM per programmare e controllare altri strumenti digitali, come API, motori di ricerca e persino altri sistemi di intelligenza artificiale generativa (Schick et al., 2023; Mialon et al., 2023; Chase, 2022). Ciò consente l'integrazione senza soluzione di continuità di singoli componenti per una maggiore utilità, prestazioni e generalizzazione. Nel lungo periodo, questi trend suggeriscono che i LLM potrebbero essere in grado di eseguire qualsiasi compito tipicamente svolto su un computer.
I modelli di intelligenza artificiale generativa sono stati utilizzati principalmente come moduli specializzati, svolgendo compiti specifici come la generazione di immagini da brevi descrizioni o la trascrizione di testo da discorsi. Possiamo però adottare una prospettiva più ampia e vedere i LLM come componenti funzionali per la costruzione di nuovi strumenti. Sebbene la costruzione di questi strumenti e la loro integrazione in sistemi completi richieda tempo e una significativa riconfigurazione dei processi esistenti nell'economia, già oggi si osservano tendenze emergenti di adozione. Nonostante le loro limitazioni, i LLM stanno diventando sempre più integrati in applicazioni specializzate in aree come l'assistenza alla scrittura, la programmazione e la ricerca legale, aprendo la strada per un'adozione più diffusa dei GPT.
È importante considerare questi sistemi più articolati e completi che integrano le funzionalità dei LLM anche perché i soli modelli LLM generalisti pronti all›uso possono continuare a essere poco affidabili per vari compiti a causa di problemi come l'inaccuratezza dei fatti, i *bias* intrinseci, le preoccupazioni per la privacy e i rischi di disinformazione (Abid et al., 2021; Schramowski et al., 2022; OpenAI, 2023a). I sistemi costruiti a supporto di processi verticali specifici possono contribuire a risolvere queste limitazioni incorporando competenze specifiche di dominio.
È interessante considerare che potrà emergere un circolo virtuoso, quando i LLM supereranno una certa soglia di prestazioni e potranno contribuire essi stessi alla costruzione degli strumenti che migliorano la loro utilità e usabilità in vari contesti. Ciò potrebbe ridurre il costo e le competenze ingegneristiche necessarie per creare tali strumenti, accelerando ulteriormente l›adozione e l›integrazione dei LLM (Chen et al., 2021; Peng et al., 2023). I LLM possono anche diventare strumenti preziosi nello sviluppo di modelli di *machine learning*, servendo come assistenti allo sviluppo e coding per ricercatori, come servizi di annotazione dei dati o generatori di dati sintetici. Man mano che i LLM migliorano nel tempo e si allineano meglio alle preferenze dell'utente, possiamo prevedere un continuo miglioramento delle prestazioni.

3.1 Fermento nell'industria del software e prime applicazioni

In virtù delle interessanti caratteristiche dei LLM e delle loro possibili applicazioni, assistiamo al proliferare di nuovi modelli annunciati e resi disponibili da diverse aziende e istituzioni. Sicuramente il più noto è GPT-3, realizzato da OpenAI, che fin dai suoi esordi nel 2020 ha sorpreso anche i non addetti ai lavori (famoso è stato l'articolo pubblicato sul "The Guardian" l'8 settembre 2020 scritto interamente da GPT-3[11]). La notorietà di GPT-3 è diventata ancora maggiore con una diffusione quasi dirompente, da fine novembre 2022, quando Ope-

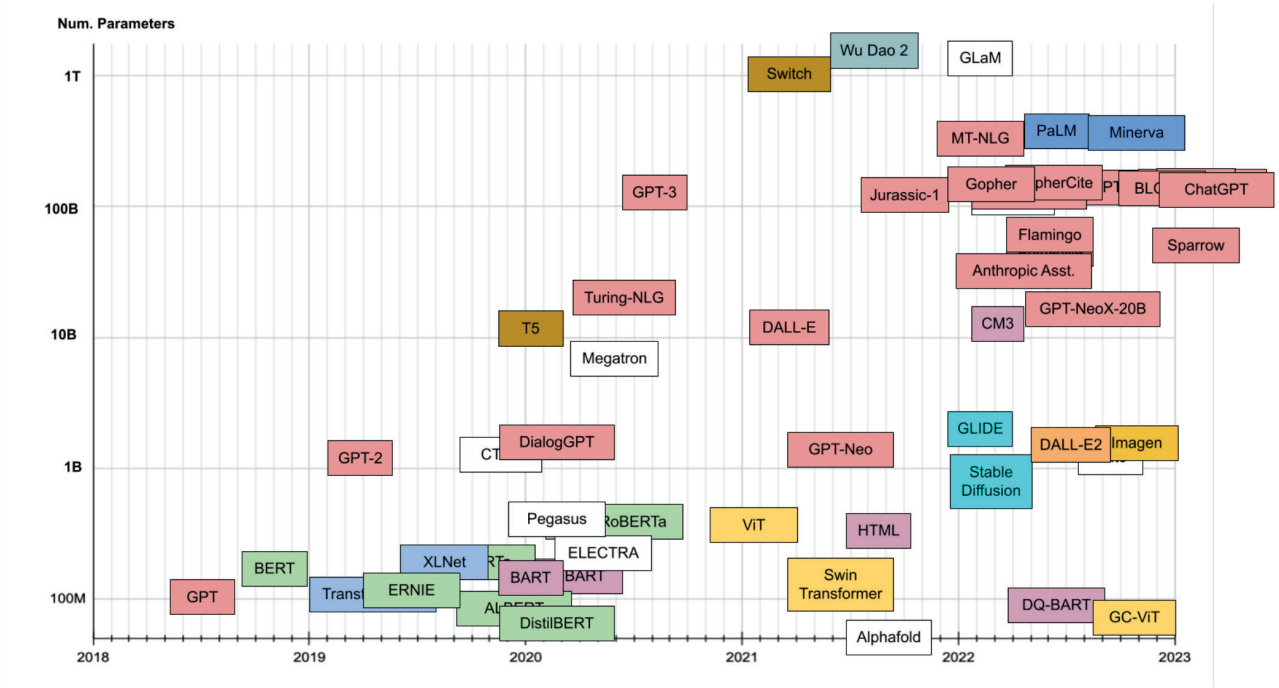

Figura 2 - Principali modelli basati su architettura a Transformer. Fonte: Xavier Amatriain, *Transformer models: an introduction and catalog*, 2023, https://arxiv.org/abs/2302.07730

nAI ha reso pubblicamente accessibile in un ambiente sperimentale ChatGPT, un chatbot basato su GPT-3 che ha consentito a milioni di persone nel mondo di interagire via chat con il modello linguistico, sperimentando in prima persona i limiti e le potenzialità di questi strumenti.

Da allora quasi ogni settimana viene annunciato un nuovo LLM o una nuova versione. Qui di seguito vengono segnalati alcuni tra i più recenti e importanti:

- GPT-4, l'ultima evoluzione dei LLMs di OpenAI;
- LaMDA, su cui si basa il recentissimo sistema di ricerca conversazionale di Google BARD;
- Google Flan-T5;
- LLaMA, recentemente annunciato da Mark Zuckerber per Meta;
- Claude, assistente AI dell'azienda Antrophic;
- Cohere, realizzato da una start up canadese;
- Forefront.ai, servizio che rende disponibili diversi modelli Open Source;
- Bloom, della azienda Hugging.Face.

In Figura 2 un diagramma mostra i principali modelli basati su architettura a Transformer indicando sulle ascisse l'anno di introduzione e sulle ordinate la dimensione di ciascun modello.

Nel contesto attuale gli sforzi nella realizzazione di LLM vedono come attori principali le grandi imprese del software come Meta, Microsoft (OpenAI) e Google che dispongono di ingenti risorse per finanziare i costi di addestramento di modelli molto sofisticati, alcune startup focalizzate nel declinare i LLM in compiti specifici e una estesa comunità che contribuisce alla definizione e manutenzione di modelli Open Source. Il contributo aperto alla comunità di addetti ai lavori è molto importante per garantire il progresso della ricerca, la democratizzazione nell'utilizzo, la possibilità di adozione e di personalizzazione di questa importante tecnologia in modo il più possibile diffuso e trasversale per contenere l'accentramento delle competenze e delle capacità di innovazione in capo a pochi soggetti.

A inizio 2023 Nat Friedman ha realizzato nat.dev,[12] un piccolo strumento online in cui si possono confrontare le caratteristiche e prestazioni di molti diversi LLMs, sia Open Source sia proprietari. È uno strumento utile per osservare come i diversi gruppi di ricerca declinano le capacità dei loro modelli e come questi generino le rispettive risposte a fronte del nostro medesimo *prompt*. È abbastanza immediato verificare come GPT-3 e GPT-4 offrono risultati migliori in tutti i compiti.

Oltre allo sviluppo dei modelli, nel panorama delle applicazioni rivolte al pubblico più vasto si può notare una crescente diffusione di funzionalità basate sui modelli di linguaggio. I principali produttori di software, infatti, stanno iniziando ad arricchire i loro prodotti integrando le capacità offerte dai LLM per semplificare le attività di scrittura, per la gestione dei rapporti con i clienti o per gestire le attività di vendita.

La crescente attenzione e importanza che questa tecnologia rappresenta oggi si può anche determinare osservando come Microsoft e Google stiano rivaleggian-

do per prepararsi a fornire i migliori servizi potenziati dall'AI per i loro utenti. Tutte le applicazioni di *Office Automation* di uso quotidiano verranno arricchite di nuove funzionalità basate su LLMs e anche il modo con cui eseguiamo le ricerche sul web è in profonda trasformazione supportata proprio da LLM sempre più performanti. Già ora possiamo sperimentare qualche anteprima con le nuove versioni di Microsoft Office 365 Copilot, Microsoft Bing e Google Bard.[13]
In questo scenario cerchiamo di descrivere il fenomeno GPT e il suo posizionamento nel contesto dei LLM.

4. OpenAI GPT: un LLM all'avanguardia

Generative Pre-trained Transformer 3 (GPT-3) è un modello di intelligenza artificiale sviluppato da OpenAI che ha rivoluzionato il campo del *Natural Language Processing*. GPT-3 è il terzo modello della serie GPT ed è incredibilmente potente grazie alla sua capacità di generare testo coerente e contestualmente corretto. Questo modello ha dimostrato capacità sorprendenti nel comprendere e generare testo in diverse lingue e in vari contesti, rendendolo uno strumento versatile e adatto a molteplici applicazioni.
GPT-3, su cui esiste una abbondante letteratura scientifica di riferimento,[14] sfrutta un'architettura chiamata Transformer, già richiamata in precedenza, originariamente introdotta nel 2017 da Vaswani et al. Il Transformer è un tipo di rete neurale che si basa su meccanismi di attenzione per processare sequenze di dati di lunghezza variabile. L'architettura Transformer utilizza principalmente due componenti chiave: l'attenzione *multi-head* e il posizionamento degli input nella sequenza. L'attenzione *multi-head* consente al modello di attribuire importanza a diverse parole nel contesto, mentre il posizionamento degli input aiuta il modello a comprendere l'ordine delle parole nella sequenza. Questi due meccanismi permettono a GPT-3 di catturare relazioni complesse tra parole e contesti.
GPT-3 viene pre-allenato su un vasto corpus di testo non supervisionato, che include siti web, libri, articoli e altri tipi di contenuti. Il modello viene allenato per minimizzare l'errore sulla probabilità di generare correttamente la prossima parola in una sequenza, data la sequenza di parole precedenti.
Una volta completato il pre-allenamento, GPT-3 può essere adattato a specifici compiti di elaborazione del linguaggio naturale (NLP) attraverso un processo chiamato *fine-tuning*. Durante il *fine-tuning*, GPT-3 viene allenato su un insieme di dati etichettati specifico per il compito, con l'obiettivo di minimizzare l'errore per quel compito. Il *fine-tuning* permette a GPT-3 di acquisire conoscenze specifiche del dominio e di adattarsi alle esigenze di specifiche applicazioni (Brown et al., 2020).

4.1 ChatGPT, l'esperienza collettiva

A fine novembre 2022 OpenAI ha reso pubblicamente accessibile ChatGPT, un ambiente di test dove milioni di persone possono testare le capacità di GPT-3 attraverso una semplice interfaccia conversazionale.
ChatGPT introduce almeno due importanti novità. La prima è senz'altro la modalità di interazione semplice e immediata: possiamo chiedere al modello GPT-3 di eseguire ogni compito semplicemente dialogando e al contempo fornire una nostra valutazione della qualità del risultato ottenuto. La seconda importante novità è il *fine-tuning* che OpenAI ha adottato per rendere GPT-3 in grado di funzionare e produrre risposte adeguate a una interazione conversazionale utilizzando una tecnica di *machine learning* chiamata *Reinforcement learning from human feedback* (RFHF).
La semplicità di interazione e l'ampia diffusione di ChatGPT hanno contribuito a determinare la curiosità e l'interesse per i LLM e di fornire a OpenAI numerosi feedback e commenti sia critici che di grande apprezzamento che sono stati utilizzati per progredire lo sviluppo delle future versioni.

4.2 Perché GPT è rivoluzionario. La lingua come modello di ragionamento?

Come abbiamo visto il progresso più rilevante della ricerca nell'AI degli ultimi anni è stata l'evoluzione del *Natural Language Processing* raggiunta dai grandi modelli di linguaggio, come GPT-3.
Al di là degli aspetti più propriamente tecnici e scientifici molto complessi che sottendono all'esistenza di questi modelli di linguaggio e che a loro volta si fondano su concetti e nozioni che richiedono anni di studio per poter essere compresi, perché i LLMs come GPT-3 sono da ritenere così importanti?
L'architettura di rete a *Transformer* introdotta nel famoso articolo *Attention is all you need* da un team di ricercatori di Google Brain nel 2017[15] consente agli algoritmi di *machine learning* basati sulle reti neurali che sfruttano tale architettura di imparare e riconoscere le regole del linguaggio naturale semplicemente osservando enormi quantitativi di testo senza la necessità di fornire dei dati di addestramento specificamente preparati per lo scopo. Questo è stato l'inizio della rivoluzione. Possiamo affermare che l'introduzione dei Transformer ha dato l'avvio a un cambio di paradigma nell'affrontare e

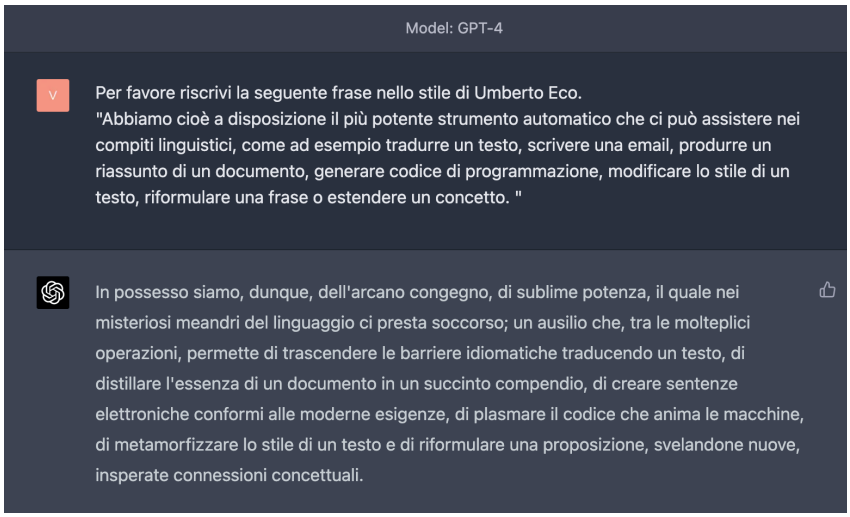

Figura 3 - Esempio in cui GPT-4 riscrive una frase modificandone lo stile

risolvere una classe di problemi di *machine learning*. Un primo risultato sorprendente dei LLMs addestrati utilizzando questa tecnica e le sue molteplici derivazioni e affinamenti è appunto la loro abilità nel rappresentare il linguaggio naturale senza un addestramento specifico.
Tradizionalmente per ogni specifico compito di analisi del testo scritto, i relativi modelli di *machine learning* venivano addestrati preparando costosi insiemi di esempi per ogni passo del processo di analisi. Nonostante i discreti risultati che otteniamo, questo approccio ha sempre mostrato due evidenti limiti: da un lato il costo altissimo di etichettare a mano ogni singola parola all'interno di un corpus di documenti necessario per l'addestramento e la conseguente limitazione in termini di varietà di esempi possibili e di rappresentazione delle diverse lingue (per ciascuna lingua decine di corpus con milioni di documenti dovevano essere predisposti con una etichettatura manuale); dell'altro lato, di conseguenza, prestazioni appena discrete e un progresso lento.
Ora, semplificando, abbiamo la possibilità di sottoporre agli algoritmi di addestramento un mare sconfinato di testi in tutte le lingue e linguaggi, e questi derivano in autonomia le regole del funzionamento del linguaggio umano, imparano cosa sono le parole, la loro morfologia, la sintassi, la grammatica e la semantica. Anzi, ora con un sistema evoluto come GPT-3, possiamo chiedere al modello così ottenuto di eseguire l'analisi di una frase in (quasi) qualunque lingua, di eseguire ad esempio il *part-of-speech* o *POS tagging*[16] o derivarne l'albero sintattico e il LLMs, senza nessun addestramento specifico, genererà un output con una qualità comparabile e se non superiore a quella di modelli e algoritmi specificamente costruiti e addestrati per farlo.
Ecco un piccolo esempio dello stupore che questi modelli, e GPT-3 in particolare, hanno suscitato dapprima nella comunità scientifica degli addetti ai lavori. La loro capacità di eseguire tutti i diversi compiti tipici della linguistica computazionale e del *Natural Language Processing*, come ad esempio la sintesi di testi lunghi o la traduzione tra due lingue qualunque o da lingue naturali verso linguaggi di programmazione, con una qualità analoga o superiore ad altri modelli costruiti appositamente ciascuno per lo specifico scopo e perdipiù senza aver predisposto dati di training specifico per quei compiti.
Abbiamo a disposizione un potente strumento automatico che ci può assistere nei compiti linguistici, come ad esempio tradurre un testo, scrivere una e-mail, produrre un riassunto di un documento, generare un codice di programmazione, modificare lo stile di un testo, riformulare una frase o estendere un concetto. In Figura 3 si propone un gioco in cui GPT-4 tenta di riformulare quest'ultima frase utilizzando uno stile che si rifà alla scrittura di Umberto Eco.
Disponiamo quindi di un ottimo motore linguistico in grado di eseguire compiti per i quali non è stato specificatamente addestrato e di stupirci con abilità inattese, abilitando una seria ampia di applicazioni che potranno interagire con gli utenti attraverso semplici istruzioni in linguaggio naturale, e assistere noi utilizzatori in alcune attività più routinarie, che in ambiente Microsoft sono state definite quelle di un *copilota*.[17]
Ogni giorno scopriamo nuove abilità di GPT in contesti diversi, dalle attività di linguistica computazionale, dalla generazione di testo con qualità simile al testo generato da un essere umano, alla risoluzione di problemi di logica, di finanza, di informatica e così via. Le recenti versioni di LLM che si susseguono mostrano notevoli capacità in una varietà ampia di domini e attività mettendo in discussione la nostra comprensione dell'apprendimento dei processi cognitivi.
Sembrerebbe quindi che questi modelli più sofisticati, venendo esposti a quantitativi enormi di testi, nell'apprendere il significato delle parole e le regole del linguaggio naturale, abbiano anche in qualche modo appreso alcune abilità di astrazione e di ragionamento rivelando forse come noi umani abbiamo trasferito nella lingua alcuni dei nostri processi logici e di ragionamento e che questi siano stati in parte desunti e trasferiti nelle sinapsi di questi nuovi modelli del linguaggio.

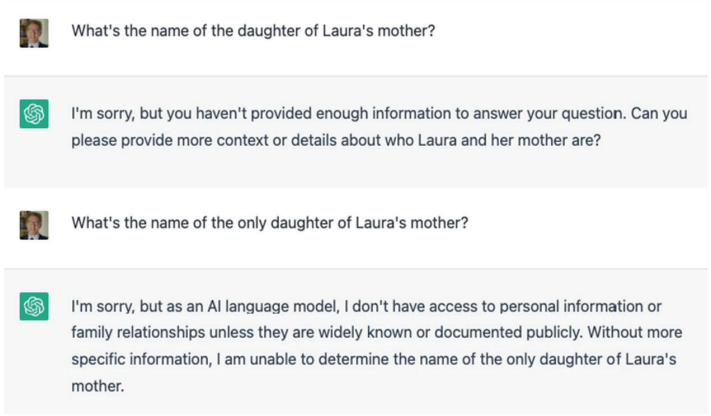

Figura 4 - Esempio proposto nell'articolo di Luciano Floridi

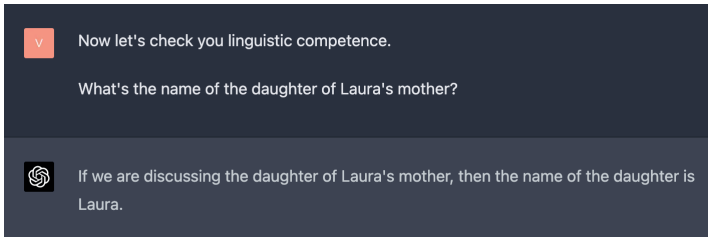

Figura 4a - Lo stesso esempio utilizzando ChatGPT con il modello GPT-4 e la frase proposta in lingua inglese

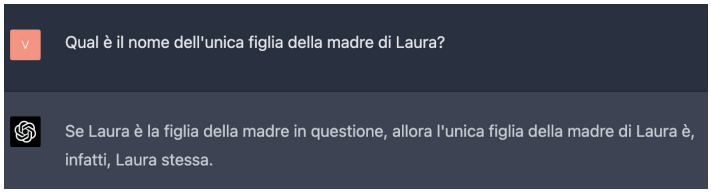

Figura 4b - Lo stesso esempio utilizzando ChatGPT con il modello GPT-4 e la frase proposta in lingua italiana

4.3 GPT-4: caratteristiche e potenzialità

GPT-4 (OpenAI, 2023b) estende ulteriormente le capacità e le potenzialità dei modelli di linguaggio: le migliorie rispetto al precedente GPT-3 riguardano la qualità delle risposte generate, le capacità di astrazione e di ragionamento. GPT-4 ha dimostrato prestazioni di livello umano nella maggior parte degli esami accademici e professionali su cui è stato misurato.
Notevole, ad esempio, il risultato ottenuto nell'esame *Uniform Bar Examination* (UBE) che ha superato con un punteggio che si posiziona nel 10% dei primi di tutti i punteggi ottenuti dai partecipanti a quell'esame.[18]
Come ampiamente documentato nel report associato alla pubblicazione di GPT-4, la nuova versione supera GPT-3 in tutti i *benchmark* con cui i modelli vengono valutati. In decine di diversi test accademici e di accesso alle professioni a cui è stato sottoposto, i risultati ottenuti da GPT-4 si posizionano tra i migliori in assoluto.
GPT-4 è stato testato rispetto tutti i criteri di valutazione progettati per misurare le prestazioni dei *Language Models* e ha superato in modo considerevole tutti i modelli esistenti ridefinendo lo stato dell'arte in svariati campi di applicazione.
GPT-4 non si distingue per l'ampiezza della sua conoscenza, sempre limitata a settembre 2021, ma per le sue capacità di elaborazione del linguaggio naturale, a cui si aggiunge la capacità nuova, che non era presente nelle versioni precedenti, di interpretare non solo i testi ma anche le immagini, aprendo a nuovi scenari di applicazione.
Al momento della stesura dell'articolo GPT-4 è disponibile attraverso ChatGPT solo per gli utenti che hanno sottoscritto un abbonamento a pagamento, oppure via API transitando per una lista di attesa. Personalmente lo sto utilizzando ed eseguendo i primi esperimenti che confermano la maggiore qualità delle risposte che ottengo e la miglior qualità di ragionamento e di problem solving.
Di seguito un piccolo esempio suggerito da un recente articolo di Luciano Floridi, in cui sottoponeva a GPT-3.5 una semplice domanda di comprensione del testo come riportato in Figura 4.[19]
Ho provato a porre la stessa domanda utilizzando il nuovo modello GPT-4 ottenendo la risposta corretta (Figure 4a e 4b).

4.4 GPT-4 è un passo verso la AGI?

Una differenza fondamentale tra intelligenza artificiale e intelligenza artificiale generale (AGI, dall'inglese *Artificial General Intelligence*) è che l'AI è ristretta all'esecuzione di un compito specifico (riconoscere il volto di una persona in un'immagine, giocare a scacchi, rimuovere lo spam della casella di posta ecc.), mentre un'ipotetica AGI dovrebbe poter affrontare problemi di qualsiasi natura come un essere umano.
L'intelligenza artificiale generale si riferisce quindi a un'AI altamente autonoma che possiede la capacità di apprendere, comprendere, adattarsi e applicare la propria intelligenza in un'ampia varietà di compiti e contesti, in modo simile a quanto fa un essere umano.[20]
Sebbene non ci siano al momento analisi o dati che sostengono che GPT-4 sia un esempio di intelligenza artificiale generale esistono tuttavia numerosi elementi per ritenere, tenendo conto delle affermazioni di Bill Gates

riportate in apertura, che siamo all'alba di qualcosa di nuovo, che riguarda tutti e che ci richiede un coinvolgimento diretto per poter contribuire nell'immaginare i prossimi passi e le migliori ricadute e applicazioni.
Sam Altman, CEO di OpenAI, in una recente intervista ha così definito GPT-4:

> Che cos'è GPT? È un sistema a cui guarderemo indietro dicendo che era una Intelligenza Artificiale molto giovane, era lenta e difettosa, molte cose non le riuscivano bene, ma lo stesso pensiamo oggi dei primissimi computer che hanno definito la strada verso gli oggetti che oggi sono diventati molto importanti nelle nostre vite anche se hanno richiesto decenni per evolvere.[21]

Un articolo recentemente pubblicato dai ricercatori di Microsoft Research[22] riporta alcune evidenze che GPT-4 mostra molte tracce di intelligenza secondo la definizione che ne diede nel 1994 un gruppo di 52 psicologi che sottoscrissero un editoriale fondativo della scienza dell'intelligenza.[23] Il gruppo di lavoro definì l'intelligenza come una capacità mentale molto generale che, tra le altre cose, coinvolge la capacità di ragionare, pianificare, risolvere problemi, pensare in modo astratto, comprendere idee complesse e imparare rapidamente dall'esperienza. Questa definizione implica che l'intelligenza comprenda un'ampia gamma di abilità e competenze cognitive.
Secondo i ricercatori, GPT-4, pur essendo un modello di linguaggio, dimostra considerevoli capacità in una varietà di domini e attività, incluse astrazione, visione, programmazione, matematica, medicina, legge, comprensione delle emozioni umane e molto altro.
Nel corso dello studio i ricercatori raccolgono le evidenze di queste capacità eseguendo decine di test molto vari ed estesi andando oltre i benchmark classicamente usati per misurare le prestazioni di un sistema di AI, concludendo che GPT-4 sembra mostrare qualche traccia di AGI.
Nel prossimo futuro vedremo certamente molta attenzione da parte della comunità scientifica nell'interpretare i risultati ottenuti dai numerosi LLMs, analizzarne i limiti e le debolezze e nel provare a sviluppare nuove versioni migliori e più affidabili. I risultati futuri naturalmente sono incerti ma ci pervade la sensazione che potremo assistere a un rapido e continuo progresso delle capacità dei prossimi modelli del linguaggio.

5. Allucinazioni di GPT: affidabilità e accuratezza delle informazioni

Il fatto che GPT-3 possa fornire risposte errate a semplici domande solleva dubbi sulla sua intelligenza e sulle sue capacità cognitive.
Le cosiddette "allucinazioni" di GPT-3 sono una caratteristica peculiare del modello, che talvolta fornisce risposte errate o inventa informazioni con un tono assertivo. Questo fenomeno solleva preoccupazioni riguardo l'affidabilità e l'utilizzo di GPT-3, soprattutto nei contesti dove l'accuratezza e l'affidabilità delle informazioni sono fondamentali.
Nonostante i progressi compiuti nell'addestramento e nella configurazione dei modelli, i LLM, pur essendo in grado di generare testi coerenti e plausibili, possono generare informazioni inesatte, fuorvianti o obsolete[24] a causa di diverse ragioni, tra cui:

1. *Bias* nei dati di addestramento: i modelli apprendono dai dati a loro disposizione, che possono contenere informazioni sbagliate, fuorvianti o parziali. Questo può portare a risposte inesatte o distorte;
2. *Overfitting*: i LLM possono memorizzare informazioni specifiche dai dati di addestramento, il che può portare a risposte obsolete o non più valide nel contesto attuale;
3. Ambiguità nella domanda: le domande degli utenti possono essere ambigue o poco chiare, il che può portare il modello a generare risposte inesatte o fuorvianti.

Per interpretare i risultati di GPT-3, è *essenziale* tenere a mente che il modello è un generatore di testo e non un esperto di dominio. Pertanto, le risposte fornite dal modello devono essere valutate criticamente e confrontate con altre fonti autorevoli prima di considerarle attendibili. Inoltre, è possibile utilizzare tecniche di mitigazione, come il controllo dell'output del modello o l'introduzione di un feedback umano, per ridurre il rischio di allucinazioni.[25]
Le ragioni sopra citate, per quanto non esaustive, ci conducono a una importante riflessione.
I principali limiti e problematiche che vengono più comunemente riscontrati nei modelli sofisticati come GPT-3 o GPT-4 sono per la maggior parte riconducibili alle informazioni contenute nei testi generati in risposta a qualche nostro stimolo, informazioni che talvolta possono essere false o incongruenti. La valutazione delle capacità del modello basata sulla verità delle informazioni restituite è però una valutazione parziale e limitata a un aspetto che originariamente è considerato secondario.
Come abbiamo visto i LLM esistono e sono stati addestrati per svolgere compiti tipici della trattazione del linguaggio naturale e non per interpretare il ruolo di oracoli onniscienti o di esperto di qualche dominio.
È importante quindi distinguere due tipologie di abilità:

1. Capacità linguistiche, di astrazione e di generazione di testo coerente;
2. Conoscenza di dominio.

Se valutiamo un LLM come GPT-3 e GPT-4 sulla base dei criteri che afferiscono alla prima abilità, è dimostrato che i risultati sono sorprendentemente positivi, forse non sempre perfetti, ma certamente molto efficaci e sorprendenti.
Per quanto riguarda invece le abilità che necessitano anche di generare risposte accurate, coerenti, il più possibile senza *bias,* è necessario ulteriore lavoro. È fondamentale sviluppare meccanismi di controllo e validazione delle risposte dei modelli, integrando l›esperienza specifica del dominio e il coinvolgimento umano nel processo decisionale.
Entro questi limiti è probabile che potremmo disporre di molti modelli che, partendo da grandi modelli pre-addestrati, verranno raffinati (*fine-tuning*) su conoscenze specifiche per diversi domini di conoscenza. È un ambito di lavoro molto interessante che coinvolge diverse professionalità e discipline e vede esperti di *machine learning* lavorare insieme a esperti di dominio per raffinare e migliorare sia la capacità del modello di interagire con gli utenti sia le sue competenze per ridurre così le aberrazioni. Un ambito di applicazione di grande interesse, che vedrà concentrarsi gli sforzi di una auspicabilmente vasta comunità, è l'utilizzo di futuri LLMs come strumenti di supporto nell'ambito dell'insegnamento.
Nel frattempo, sia OpenAI sia Google stanno iniziando a rendere disponibile la possibilità per i loro LLMs (GPT-x e LaMBA) di reperire dati e informazioni aggiornate per abilitare la generazione di risposte non limitate sui dati acquisiti in fase di training ma integrate da nozioni specifiche rispetto alla domanda posta. I modelli hanno così la possibilità di estendere la conoscenza necessaria per svolgere il compito richiesto attingendo in tempo reale a fonti esterne (ad esempio notizie sul web, gli orari dei collegamenti aerei, il calendario degli eventi, l'archivio dei documenti o dei contatti) mitigando le allucinazioni talvolta contenute nei testi generati e aprendo la strada a nuove possibili applicazioni.

6. Conclusioni

Stiamo attraversando un periodo peculiare di grande fermento, i recenti modelli di linguaggio riscuotono notevole interesse per le loro capacità inattese ma sono ancora strumenti immaturi che richiedono ulteriori sforzi per una loro incerta prossima evoluzione.
Il procedere della loro evoluzione potrebbe condurre a capacità particolarmente rilevanti che potrebbero ridefinire come interagiamo con le macchine e come affrontiamo certi problemi, dall'educazione alla creazione di contenuti, dalla ricerca scientifica alle strategie economiche.
Seppur con poche certezze riguardo alle effettive future capacità, è importante procedere parallelamente nelle attività necessarie a garantire che le prossime AI siano allineate alle intenzioni dei loro progettisti e agli obiettivi di benessere per la collettività. Per questo motivo tutti noi siamo chiamati a riflettere e dibattere su queste tematiche contribuendo al progresso della tecnologia nella direzione che riteniamo più utile e benefica per i nostri obiettivi.

BIBLIOGRAFIA

Abid et al., 2021 = Abubakar Abid, Maheen Farooqi, James Zou, *Persistent anti-muslim bias in large language models,* "Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society" (2021), p. 298-306, https://doi.org/10.48550/arXiv.2101.05783.

Bai et al., 2022 = Yuntao Bai et al., *Training a helpful and harmless assistant with reinforcement learning from human feedback,* arXiv preprint (2022), arXiv:2204.05862, https://doi.org/10.48550/arXiv.2204.05862.

Brown et al., 2020 = Tom Brown et al., *Language models are few-shot learners,* "Advances in neural information processing systems", 33 (2020), p. 1877-1901.

Chase, 2022 = Chain Lang, https://github.com/hwchase17/langchain.

Chen et al., 2021 = Mark Chen et al., *Evaluating large language models trained on code,* arXiv preprint (2021), arXiv:2107.03374, https://doi.org/10.48550/arXiv.2107.03374.

Hernandez et al., 2021 = Danny Hernandez et al., *Scaling laws for transfer,* arXiv preprint (2021), arXiv:2102.01293, https://doi.org/10.48550/arXiv.2102.01293.

Kaplan et al., 2020 = Jared Kaplan et al., *Scaling laws for neural language models,* arXiv preprint (2020), arXiv:2001.08361, https://doi.org/10.48550/arXiv.2001.08361.

Mialon et al., 2023 = Grégoire Mialon et al., *Augmented language models: a survey,* arXiv preprint (2023), arXiv:2302.07842, https://doi.org/10.48550/arXiv.2302.07842.

OpenAI, 2023a = *Gpt-4 system card,* Technical report, OpenAI, https://cdn.openai.com/papers/gpt-4-system-card.pdf.

OpenAI, 2023b = *Gpt-4 technical report,* Technical report, OpenAI, https://doi.org/10.48550/arXiv.2303.08774.

Ouyang et al., 2022 = Long Ouyang et al., *Training language models to follow instructions with human feedback,* "Advances in Neural Information Processing Systems", 35 (2022), https://doi.org/10.48550/arXiv.2203.02155.

Peng et al., 2023 = Sida Peng, et al., *The impact of ai on developer productivity: Evidence from github copilot,* arXiv preprint arXiv:2302.06590 (2023), https://doi.org/10.48550/arXiv.2302.06590.

Radford et al., 2019 = Alec Radford et al., *Language models are unsupervised multitask learners,* "OpenAI blog", 8 (2019), 1, p. 9, https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf.

Schick et al., 2023 = Timo Schick et al., *Toolformer: Language models can teach themselves to use tools*, arXiv preprint (2023), arXiv:2302.04761, https://doi.org/10.48550/arXiv.2302.04761.
Schramowski et al., 2022 = Patrick Schramowski et al., *Large pre-trained language models contain human-like biases of what is right and wrong to do*, "Nature Machine Intelligence", 4 (2022), 3, p. 258-268, https://www.nature.com/articles/s42256-022-00458-8.
Thoppilan et al., 2022 = Romal Thoppilan et al., *Lamda: Language models for dialog applications*, arXiv preprint (2022), arXiv:2201.08239, https://doi.org/10.48550/arXiv.2201.08239.
Vaswani et al., 2017 = Ashish Vaswani et al., *Attention is all you need*, "Advances in neural information processing systems" (2017), https://doi.org/10.48550/arXiv.1706.03762.

NOTE

[1] Bill Gates, *The age of AI has begun: artificial intelligence is as revolutionary as mobile phones and the Internet*, "GatesNote. The blog of Bill Gates", March 21, 2023, https://www.gatesnotes.com/The-Age-of-AI-Has-Begun.

[2] *Ibidem*. Traduzione dell'autore.

[3] Tom Brown et al., *Language models are few-shot learners*, "Advances in neural information processing systems", 33 (2020), p. 1877-1901.

[4] Open AI 2023, *Gpt-4 technical report*, https://cdn.openai.com/papers/gpt-4.pdf.

[5] Ian Goodfellow et al., *Generative adversarial nets*, in *Advances in neural information processing systems*, 2014, p. 2672-2680.

[6] Ian Goodfellow et al., *Generative adversarial networks*, arXiv:1406.2661, https://arxiv.org/abs/1406.2661;
Diederik P. Kingma, Max Welling, *Auto-encoding variational bayes*, arXiv:1312.6114, https://arxiv.org/abs/1312.6114; Aaron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu, *Pixel recurrent neural networks*, arXiv:1601.06759, https://arxiv.org/abs/1601.06759; A. Radford et al., *DALL-E: Creating images from text*, "OpenAI Blog", 2021, https://openai.com/blog/dall-e/; Jonathan Ho, Ajay Jain, Pieter Abbeel, *Denoising Diffusion Probabilistic Models*, 2020, https://arxiv.org/abs/2006.11239.

[7] https://google-research.github.io/seanet/musiclm/examples.

[8] https://musi-co.com/.

[9] https://www.riffusion.com/.

[10] T. B. Brown et al., *Language models are few-shot learners*, "Advances in Neural Information Processing Systems", 33 (2020), p. 1877-1901; A. Elgammal et al., *CAN: Creative adversarial networks generating "art" by learning about styles and deviating from style norms*, 2017, arXiv preprint arXiv:1706.07068; I. Goodfellow et al., *Generative adversarial nets*, "Advances in neural information processing systems", 2014, p. 2672-2680; A. G. Huang, I. Sutskever, *WaveNet: A generative model for raw audio*, 2017, arXiv preprint arXiv:1609.03499; T. Karras et al., *Progressive growing of GANs for improved quality, stability, and variation*, in *International Conference on Learning Representations (ICLR)*, 2018; A. Vaswani et al., *Attention is all you need*, in *Advances in neural information processing systems*, 2017, p. 5998-6008.

[11] https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

[12] https://nat.dev/compare.

[13] https://www.bing.com/new; https://bard.google.com; https://www.microsoft.com/en-us/microsoft-365/blog/2023/03/16/introducing-microsoft-365-copilot-a-whole-new-way-to-work.

[14] J. L. Ba, J. R. Kiros, G. E. Hinton, *Layer Normalization*, 2016, arXiv:1607.06450; T. Brown et al., *Language models are few-shot learners*, "Advances in neural information processing systems", 33, (2020) p .1877-1901; N. Srivastava et al., *Dropout: A simple way to prevent neural networks from overfitting*, "The Journal of Machine Learning Research", 15 (2014), 1, p. 1929-1958; A. Vaswani et al., *Attention is all you need*, cit., p. 5998-6008.

[15] A. Vaswani, *Attention is all you need*, cit., p. 5998-6008.

[16] Il *POS tagging* consiste nella marcatura di parole di un corpus testuale per identificarne le funzioni grammaticali: cfr. https://en.wikipedia.org/wiki/Part-of-speech_tagging.

[17] Jaret Spataro, *Introducing Microsoft 365 Copilot – Your copilot for work*, Microsoft Official Blog, march 2023, https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work.

[18] L'UBE è un esame standardizzato creato dalla National Conference of Bar Examiners (NCBE). È progettato negli Stati Uniti per verificare le conoscenze e le abilità che ogni avvocato deve avere prima di ottenere la licenza per esercitare la professione forense, https://en.wikipedia.org/wiki/Bar_examination_in_the_United_States.

[19] L. Floridi, *AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models*, "Philosophy and Technology", 36 (2023), 15, https://doi.org/10.1007/s13347-023-00621-y.

[20] B. Goertzel, C. Pennachin, *Artificial General Intelligence*, Berlino, Springer, 2007; S. Legg, M. Hutter, *A collection of definitions of intelligence*, "Frontiers in Artificial Intelligence and applications" (2007), 57, p. 17.

[21] Sam Altman, *Sam Altman: OpenAI CEO on GPT-4, ChatGPT, and the Future of AI*, "Lex Fridman podcast" (2023), 367.

[22] Sebastien Bubeck et al., *Sparks of Artificial General Intelligence: Early experiments with GPT-4*, arXiv:2303.12712, 2023, https://arxiv.org/abs/2303.12712.

[23] Linda S Gottfredson, *Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography*, "Wall Street Journal",13 dicembre 1994.

[24] Abubakar Abid, et al., *Persistent anti-muslim bias in large language models*, "Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society", July 2021, p. 298-306.

[25] Jesse Dodge et al., *Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping*, arXiv preprint arXiv:2002.06305, 2020.

ABSTRACT
OpenAI has released ChatGPT to the public, sparking a discussion on the state-of-the-art advancements in recent artificial intelligence models. Over the past few months, tens of millions of people have interacted with the tool, prompting extensive contemplation on its potential, limitations, and critical issues. In the article, we aim to contextualize ChatGPT within the broader scope of Large Language Models' progress and their significance in advancing Artificial Intelligence. We encourage the widest possible involvement from everyone to ensure the most comprehensive benefits across various potential applications.