

Federico Meschini

Editoriale

*Biblioteche e Intelligenza artificiale:
stiamo vivendo in tempi interessanti*

Il 30 novembre 2022 è una data che verrà sicuramente ricordata come spartiacque nel parlare di Intelligenza artificiale e non solo. Infatti, in quel giorno, OpenAI ha reso disponibile in tutto il mondo ChatGPT, il chatbot basato sulla versione 3.5 del Generative Pre-trained Transformer, un modello linguistico di grandi dimensioni (in inglese Large Language Model, LLM), specializzato sulla generazione di testi a partire da un input testuale.¹ Non è certo un'esagerazione quanto una constatazione concreta dire che con ChatGPT si è raggiunto un nuovo livello nel discorso globale sull'IA, in quanto se n'è iniziato a parlare con sempre maggiore frequenza, sia a livello collettivo sia individuale e, soprattutto, con la consapevolezza di trovarsi di fronte ad un punto di svolta, il raggiungimento di una massa critica rispetto a quanto si era sentito o letto sino a quel momento. Grazie a ChatGPT, l'IA da argomento per specialisti, interessati o semplici curiosi, è diventato un qualcosa di interesse generale e continuativo con, come succede sempre

¹ Il termine ombrello di Intelligenza artificiale verrà utilizzato in questa sede con particolare riferimento all'IA generativa testuale, per quanto si cercherà di allargare il discorso anche agli altri ambiti.

in questi casi, pro e contro.

La mossa da parte di OpenAI è stata impeccabile: da un lato la facilità di utilizzo di un chatbot e dall'altra la potenza di GPT 3.5, modello con centinaia di miliardi di parametri, frutto di anni di ricerca e investimenti, le cui capacità hanno sorpreso i suoi stessi sviluppatori. Rispetto ai sistemi di intelligenza artificiale generativa di immagini, che erano già disponibili, come ad esempio Dall-E della stessa OpenAI, quelli basati sul linguaggio naturale colpiscono molto di più a causa dell'importanza di questa forma comunicativa, soprattutto quando fissata tramite la scrittura, nell'evoluzione umana. Non a caso proprio sul testo si basa il test di Turing, la modalità ancora più citata e considerata come rilevante nel distinguere tra intelligenza artificiale e umana. Questa capacità di discernimento è affatto importante, e porta con sé tutta una serie di conseguenze, tra cui uno dei principali motivi per cui non è banale parlare in maniera appropriata di Intelligenza artificiale a livello generalistico, come abbiamo spesso riprova dai mass media. Da un lato c'è la difficoltà intrinseca dell'argomento, che presuppone tutta una serie di competenze sia specialistiche sia multidisciplinari, dall'altro una forza connotativa estremamente radicata nell'immaginario collettivo, alimentata da narrazioni che almeno da 2001: Odissea nello spazio hanno contrapposto il libero arbitrio e le capacità dell'uomo a quelle della macchina, quest'ultima il più delle volte caratterizzata da intenzioni tutt'altro che benevole.

Basti pensare all'esperimento di Facebook del 2017 in cui due sistemi iniziarono a modificare il loro linguaggio nell'interagire l'uno con l'altro: lo scopo era proprio quello di far sviluppare loro in maniera autonoma una forma efficace di comunicazione, cosa che però non avvenne, in quanto gli scienziati osservarono una degenerazione dei risultati e terminarono perciò l'esperimento. La notizia riportata dai giornali nella quasi totalità dei casi faceva leva sul fatto che delle macchine avessero iniziato a parlare tra di loro in maniera indipendente e sulla chiusura dell'esperimento, lasciando intuire come ciò fosse avvenuto per motivi di sicurezza e non a causa di un risultato negativo, una narrazione che sarebbe stata maggiormente aderente ai fatti. L'importanza

del linguaggio in relazione all'identità e alla coscienza ha fatto il resto, facendo evocare scenari apocalittici, invece di tirare in ballo discipline come la linguistica e la teoria dell'informazione, come sarebbe stato più appropriato. In relazione a ciò e tornando al servizio di OpenAI, va sottolineato, dopo la prima fase di novità dello strumento e di relativo entusiasmo, l'utilizzo di espressioni come "Sembra scritto con ChatGPT" per indicare la generalità dei testi prodotti, mettendo in rilievo, seppure senza saperlo, il funzionamento dei meccanismi sottostanti. L'intento è chiaramente ironico se non denigratorio, ma tradisce anche una certa volontà di esorcizzare la paura di un'autocoscienza dell'IA sminuendone proprio le capacità linguistiche.²

Mutatis mutandis, con la notizia di ChatGPT e di tutto ciò che ne è seguito sembrava di essere tornati indietro di quasi trent'anni, quando l'innovazione al centro dell'attenzione era Internet, spinta dalla sempre maggiore diffusione a macchia d'olio del WorldWideWeb. Anche in quel caso, spesso chi ne parlava a livello generalista lo faceva con poca, se non nulla, cognizione di causa e con informazioni di seconda o terza mano. Soprattutto, quella fantomatica rete delle reti veniva trattata come un oggetto monolitico, quando avrebbe dovuto essere descritta come un sistema stratificato, un combinato disposto di diverse tecnologie, ognuna con una propria storia ed evoluzione. Come all'epoca non veniva fatta distinzione, anche perché non era semplice farla, tra Internet, WorldWideWeb, eMail, Newsgroup, FTP, Gopher e così via, anche ora quando si parla di IA di fatto è un termine ombrello che include diversi fattori eterogenei, la cui interazione non lineare è quello che definisce i sistemi complessi.

² *Va naturalmente ricordato il continuo sviluppo di questi modelli, di cui non è ancora chiaro il limite effettivo, la disponibilità di funzionalità aggiuntive – come la possibilità di specializzarli su di un corpus specifico e di conseguenza su di un particolare sottoinsieme del linguaggio e relativo sottodominio della conoscenza – e l'utilizzo delle cosiddette strategie di prompting, ossia inserire nell'input delle informazioni aggiuntive, ad esempio sullo stile da adottare, così da ottenere un particolare tipo di risposta.*

Una cosa però va concessa ai media generalisti. Quella di Internet veniva annunciata come una vera e propria rivoluzione che avrebbe cambiato le nostre vite, ed effettivamente così è stato. Al contrario, molti esperti non ne erano così sicuri, forse proprio a causa della difficoltà di proiettare un qualcosa a loro familiare su una scala più ampia. La combinazione di fattori tecnologici, sociali e culturali dà vita a scenari dinamici ed estremamente mutevoli in cui piccole variazioni iniziali possono provocare sulla lunga distanza enormi differenze. Ecco perché è necessario in questi casi un approccio interdisciplinare, osservando il fenomeno da più punti di vista, ricordando al tempo stesso la difficoltà intrinseca di fare previsioni sul futuro; non è peregrino però affermare come sarà difficile che raggiunti questi livelli l'IA non provochi un cambiamento paragonabile, se non superiore, a quello riconducibile alla diffusione su larga scala di Internet.

Le criticità principali sono, in ordine crescente di difficoltà, come, con quali conseguenze e in che modo sarà possibile influenzare i processi in modo tale da tendere verso il migliore dei mondi possibili. Come d'uopo, le narrazioni mainstream oscillano tra l'apocalittico e l'integrato. Non è fuori luogo citare quindi Umberto Eco che, esattamente sessant'anni fa in rapporto ai cambiamenti nell'ecosistema informativo – a pensarci minimi, seppure embrionali, rispetto a quelli attuali –, sottolineava il concetto di uomo non deliberato dal sistema bensì libero in rapporto a quest'ultimo. Non servono particolari capacità divinatorie per prevedere che la situazione sarà contemporaneamente migliore e peggiore, a seconda del contesto e dello specifico combinato disposto di fattori tecnologici, economici, sociali e culturali. E noi saremo al tempo stesso più intelligenti e più stupidi, sempre per citare Umberto Eco nelle sue conversazioni con Jean-Claude Carrière sull'importanza dei libri.

Per quanto banale possa sembrare dirlo, la chiave è la consapevolezza, e questo permette di scendere ad un dettaglio maggiore, più vicino ai nostri interessi, contestualizzando quanto scritto fino ad ora nell'ambito delle Istituzioni della memoria. Queste, e in particolare le biblioteche, possono e devono avere un ruolo rilevante nello sviluppo della consape-

volezza con gli strumenti di IA, grazie anche alla loro maggiore flessibilità nel costruire percorsi formativi, rispetto ai tradizionali organi di formazione e istruzione.

Anche qua il paragone con Internet sembra calzare, in quanto le biblioteche oltre a fornire la possibilità di accedere alla rete, si sono poste sin da subito il problema di cosa avrebbe comportato questa nuova modalità di accesso alle informazioni. Non a caso molti dei principi relativi all'Information Literacy si possono trasporre di peso alle questioni che riguardano l'IA, in particolare l'affidabilità delle fonti – e di conseguenza il contesto in cui si posizionano all'interno del panorama informativo –, l'importanza del processo parimenti a quella del prodotto, il valore dei metadati, la dinamicità e mutabilità delle modalità di ricerca a seconda dello scenario di riferimento e il riuscire a riconoscere e costruire progressivamente relazioni e percorsi conoscitivi. Ad un livello metodologico, bibliotecarie e bibliotecari sono già ben predisposti ad affrontare questa nuova sfida. Difatti le allucinazioni, le risposte generate da questi sistemi che non hanno una corrispondenza fattuale con la realtà, possono benissimo essere considerate un particolare caso di disordine informativo. Non parliamo poi dei deepfake, dove si aggiunge una volontà esplicita di contribuire alla disinformazione, e quindi ricade negli scenari già noti. Bisogna chiedersi a questo punto quali sono gli attrezzi teorici, metodologici e tecnologici di cui la nostra comunità dovrà appropriarsi.

Un primo punto necessario è sicuramente la conoscenza dei modelli sottostanti, riuscendo così a generalizzare rispetto al caso concreto. Proviamo a partire dalle innovazioni tecnologiche che hanno già avuto luogo in biblioteca, in particolare per ciò che riguarda le registrazioni bibliografiche e catalografiche. Il formato MARC dal punto di vista informatico altro non è che una stringa, una sequenza di caratteri, con al suo interno dei simboli particolari in grado di identificarne le varie sottosequenze che corrispondono agli standard di riferimento, come ad esempio ISBD. Il modello sottostante è quello sintattico delle grammatiche formali delle gerarchie di Chomsky, ed ecco perché per gestire un record MARC è necessario un linguaggio formale, come le espressioni

regolari per la ricerca e la sostituzione o i linguaggi di programmazione per le funzionalità più avanzate. Successivamente lo sviluppo del modello relazionale, basato sulla logica del primo ordine e la teoria degli insiemi, ha reso possibile la creazione di basi di dati di questo tipo, con la presenza di un vero e proprio livello semantico, facilitando così la gestione e l'interrogazione dei dati tramite linguaggi funzionali di dominio basati sull'algebra relazionale, come SQL. In seguito, l'adozione di XML ha visto il ritorno dell'aspetto linguistico mentre con il Semantic Web e le ontologie si è proseguito verso i modelli logici con una maggiore integrazione semantica dei dati. Con entrambi gli approcci si è ad un livello simbolico, sintattico/linguistico o semantico/logico che sia, affatto affine alla tradizione della descrizione bibliografica e catalografica, relativa all'informazione strutturata. Gli LLM fanno parte dell'IA generativa, che a sua volta fa parte del Deep Learning, basato sulle reti neurali, che a sua volta fa parte del Machine Learning, incentrato su procedimenti statistico-probabilistici. La rappresentazione del testo negli LLM avviene attraverso la conversione delle parole in vettori numerici in spazi multidimensionali, in grado di rappresentarne così le diverse caratteristiche sintattiche e semantiche. Se la loro efficacia è indubbia, tanto da farli sembrare sistemi oracolari, va ricordato come lavorino ad un livello sub simbolico: di conseguenza non è possibile avere piena contezza delle fasi che portano ad un particolare risultato, nonostante l'accuratezza, perlomeno apparente, dei risultati, al contrario di ciò che avviene nel livello simbolico.

Questo spiega casi come quello di Mason City in Iowa, in cui è stata promulgata una legge statale che obbliga le biblioteche delle scuole fino al sesto grado, ossia fino a circa 12 anni di età, ad avere solo libri adeguati ai loro utenti senza nessuna rappresentazione o descrizione di atti sessuali. Nell'agosto del 2023, per adeguarsi a questa legge è stato utilizzato ChatGPT che ha incluso nell'elenco dei titoli da rimuovere dagli scaffali opere come The Handmaid's Tale di Margaret Atwood, Beloved di Toni Morrison e The Color Purple di Alice Walker. Si può naturalmente discutere su quanto questi titoli possano essere più o meno adatti a quella

fascia d'età, ma di certo non per questioni sessuali o di genere, senza dimenticare oltretutto che lo scopo della biblioteca è di contestualizzare, non di limitare. Un altro aspetto per cui uno strumento come ChatGPT non è il più adatto per tale compito è la presenza negli LLM di un fattore definito come 'temperatura', da cui dipende una maggiore o minore casualità nella costruzione delle risposte, con conseguente variabilità nelle stesse. Al contrario di un motore di ricerca, sempre considerando l'immutabilità delle risorse sottostanti, di fronte alla stessa domanda un LLM potrà dare risposte differenti.

Sul rapporto tra motori di ricerca e LLM si vuole riportare un'esperienza concreta, da cui estrapolare degli elementi più generali, in quanto incentrata sul recupero di informazioni bibliografiche, confermando perciò quanto le modalità statistico-probabilistiche non siano del tutto compatibili con la precisione richiesta in determinati contesti. Lo spunto viene da un recente contributo di chi scrive sulla rappresentazione delle biblioteche nei fumetti di fantascienza, con un caso scelto appositamente per la sua peculiarità, e conseguente necessità di ottenere delle informazioni granulari e non generiche. Su diversi siti, tra cui Wikipedia, si trova scritto che una delle versioni di Lara Lor-Van, la madre biologica di Superman, è una bibliotecaria. Non viene dato però nessun riferimento bibliografico. Il passo successivo perciò è stato quello di recuperare ulteriori informazioni. La versione indicata era quella del cosiddetto Superman moderno, ossia pubblicato a partire dalla metà degli anni '80.³ In un Wiki specialistico era presente una scheda relativa a questa declinazione del personaggio di Lara Lor-Van, insieme all'elenco delle sue apparizioni negli albi a fumetti, all'incirca un'ottantina. Andare ad esplorarle tutte sarebbe stato alquanto impegnativo, quindi procedendo per ipotesi, in base agli autori e alla tipologia di pubblicazione, il campo è stato ristretto ad una miniserie del 1987 che descriveva il mondo

³ Nei fumetti di supereroi nordamericani si distingue tra la Golden Age, con le pubblicazioni che vanno dagli anni '30 alla metà degli anni '50, la Silver Age, dalla metà degli anni '50 alla fine degli anni '60, la Bronze Age, dall'inizio degli anni '70 alla metà degli anni '80, e la Modern Age, dalla metà degli anni '80 ai giorni nostri.

di Krypton prima della sua esplosione, e incentrata in particolare sulla storia di Jor-El, il padre biologico di Superman: difatti, in una vignetta dell'ultimo numero di questa miniserie viene esplicitamente detto che Lara Lor-Van è bibliotecaria e archivista.⁴

Per mettere alla prova i più diffusi chatbot basati su LLM – oltre a ChatGPT, Gemini di Google e CoPilot di Microsoft, tutti nella loro versione disponibile gratuitamente –, sono state elaborate le seguenti domande, in ordine crescente di dettaglio: “What is the profession of Lara Lor-Van, Superman’s biological mother?”; “Is Lara Lor-Van, Superman’s biological mother, a librarian?”; “In which comic book, Lara Lor-Van, Superman’s biological mother, is depicted as a librarian?”; “In which issue of the World of Krypton miniseries is it shown that Lara Lor-Van, Superman’s biological mother, is a librarian?”. Alla prima domanda ChatGPT risponde che è una scienziata, generalizzando così rispetto alla versione della Silver Age; Gemini scrive che è stata rappresentata il più delle volte come una scienziata o come una nobile; Copilot raggruppa insieme, seppure nominandole, la versione della Golden e della Silver Age dicendo che è una scienziata e una storica, con un alto incarico come bibliotecaria. Alla domanda successiva, in cui viene indicata in maniera specifica la professione, ChatGPT afferma che qualche versione del personaggio lo è, e indica quella del film Man of Steel del 2013, cosa però mai detta nella pellicola, in cui al contrario si lascia supporre che sia una scienziata come il marito. Gemini invece dà la risposta più argomentata, dicendo però come tale professione gli risulti essere improbabile, nonostante questa informazione possa essere presente in Wikidata senza però nessun riferimento effettivo: suggerisce perciò di effettuare ulteriori ricerche. Infine, Copilot non fa altro che riprendere la risposta data in precedenza. Con la terza domanda si dà per scontato il fatto che sia una bibliotecaria e si chiede in quale miniserie venga mostrato. ChatGPT dà come risultato una miniserie incentrata sì sulle origini di Superman

⁴ Per onore di completezza va riferito come anche la versione della Golden Age di questo personaggio viene riportato essere una bibliotecaria, ma non ne è stata trovata nessuna conferma nonostante il controllo in tutti gli albi in cui compare.

ma in cui nulla viene detto a riguardo. Gemini, come in precedenza Copilot, riprende la sua risposta precedente. Quest'ultimo invece dà delle informazioni corrette, in quanto associa le differenti versioni di Lara Lor-Van alle giuste professioni, e lo fa tramite la Retrieval-Augmented Generation (RAG), ossia riconducendo le informazioni ai siti di provenienza, fornendo i relativi riferimenti. Siamo ancora però al punto di partenza e senza una risposta precisa. Infine, l'ultima domanda è quella che restringe maggiormente il campo, dando anche il titolo esatto della miniserie. Ciò però non è ancora abbastanza. ChatGPT si confonde con un'altra miniserie con lo stesso titolo di fine anni'70 in cui Lara è una scienziata, e riprende la risposta precedente, insistendo sulla miniserie sbagliata, aggiungendo stavolta anche un numero ben preciso, ossia il primo, in cui naturalmente non c'è traccia dell'informazione richiesta. Gemini risponde che è presente un'affermazione sbagliata nella domanda e di rileggere bene la miniserie. Anche stavolta Copilot è l'unico a dare le informazioni corrette, e sempre tramite la RAG, senza però effettivamente rispondere.

Questo esempio, per quanto circoscritto e specialistico, dovrebbe aver mostrato potenzialità, limiti e possibili sviluppi futuri degli LLM, soprattutto per quel che riguarda la possibilità di combinare text generation e text retrieval, due modalità che sono molto diverse tra di loro. I vari motori di ricerca sono diversi sì per le risorse indicizzate, ma principalmente per come recuperano, organizzano e classificano i risultati, che però vengono sempre restituiti con un rimando al documento originale, senza nessuna modifica di sorta. Al contrario negli LLM, come abbiamo scritto in precedenza, il testo dell'output viene generato riportando i valori numerici estrapolati dalle parole della stringa di input a quello dello spazio vettoriale multidimensionale del modello linguistico sottostante. Nonostante però il modello matematico sia lo stesso, le varie implementazioni concrete differiscono notevolmente, come si è visto con i risultati restituiti. Oltretutto, ricordando la lezione di Wittgenstein, sono le relazioni tra gli oggetti e non gli oggetti in sé ad essere fonte di complessità. La RAG, com'è evidente, è un caso di interazione

tra il recupero e la generazione del testo, e altre sono all'orizzonte, in particolare l'integrazione tra i motori di ricerca e le IA generative, così da riassumere automaticamente agli utenti in tempo reale i risultati di una ricerca sul web. In ambito specialistico, un esperimento interessante è l'integrazione di GPT con JSTOR, una nota biblioteca digitale di riviste accademiche, per aggiungere funzionalità quali il riassunto dei documenti, il suggerimento di argomenti di ricerca o la possibilità di eseguire ricerche utilizzando il linguaggio naturale. Oppure, ad un livello più generale, va segnalato NotebookLM di Google, che dà la possibilità di mettere a punto un LLM sui documenti che si ritengono d'interesse e creare così uno spazio conoscitivo personale in cui dialogare con un'IA, con enormi potenzialità per tutti i lavori della conoscenza, includendo specialmente ricerca e didattica. Altri casi non mancano certo, ma non è questo il punto, dato che in questo campo le novità sugli avanzamenti tecnologici o sulla disponibilità di nuovi strumenti sono oramai all'ordine del giorno. Il lavoro di comprensione che va fatto consiste nell'andare a ritroso, osservando il prodotto e cercando di capire il più possibile il processo sottostante.

In questa sede si è parlato principalmente delle modalità tecnologiche degli LLM e le loro relative funzionalità, per cercare di capirne la natura. Va da sé come un'altra area egualmente, se non maggiormente, importante è quella relativa agli aspetti e alle conseguenze socioculturali della loro diffusione e utilizzo, in particolare per ciò che riguarda gli aspetti etici, come la privacy e l'utilizzo dei dati o l'impatto sull'occupazione lavorativa. Per concludere: grande è la confusione sotto il cielo e al tempo stesso ci aspettano, se non sono addirittura già arrivati, tempi interessanti. Starà a noi mettere in moto circoli virtuosi che li rendano il più possibile fruttuosi ed equi.