

"L'architetto e l'oracolo" di Gino Roncaglia

di Sara Faggionato

25-03-2024

Recensione a: Gino Roncaglia, *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a ChatGPT*, Laterza, Roma-Bari 2023, pp. 248, 19 euro (scheda libro)

In un quadro in cui il progresso e l'innovazione tecnologica paiono le stelle polari della nostra società, è necessario cercare di comprendere i cambiamenti più o meno radicali che investono i vari ambiti del nostro sistema sociale, a partire da quello economico sino a quello culturale. L'uomo, per natura, ha sempre avuto la necessità di progredire, andando incontro a un processo di sviluppo, che per definizione indica lo scioglimento di un viluppo, di una costrizione, quindi di un limite, sia esso naturale o teorico. Oggigiorno, l'innovazione tecnologica investe in particolare il campo del digitale, che si manifesta attraverso nuove forme di "intelligenza artificiale" (IA), definizione coniata per la prima volta da John McCarthy nel documento preparatorio del seminario svoltosi nell'estate del 1956 al Dartmouth College, ad Hanover, in cui emerge l'idea che l'informatica possa realizzare macchine dotate di un'intelligenza simile a quella umana. Questo sviluppo del digitale ha comportato - e tuttora comporta - delle trasformazioni di larga portata che interessano però anche il mondo della conoscenza e del sapere. A tale proposito, uno dei settori più influenzati è l'editoria, che è stata costretta a stare al passo con le nuove possibilità della tecnologia e con le nuove esigenze delle persone.

È proprio a partire da questo quadro che Gino Roncaglia, nel libro *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a ChatGPT* (Laterza 2023), analizza dettagliatamente il progresso digitale, che può definirsi a tutti gli effetti come la terza grande rivoluzione, e ne investiga la portata sulla società odierna e in particolare all'interno del mondo editoriale.

Il titolo del libro indica il fulcro attorno a cui ruota l'argomentazione di Roncaglia, dal momento che permette di delineare due modelli diversi di acquisizione e di produzione di conoscenze: "l'architetto", ovvero il sistema tradizionale, che è architettonico e sistematico, poiché raccoglie dati sulla base di un principio deterministico, a cui appartengono le enciclopedie e la catalogazione bibliotecaria; e "l'oracolo", come ChatGPT, che gestisce un enorme corpus di testi, suddividendoli in unità molto piccole, che possono essere singole parole o unità di parole, chiamate token (l'esempio utilizzato è dinosauro, divisibile in dino-sauro), attraverso cui si forma una matrice di numeri che ne rappresenta l'uso nel linguaggio, ovvero nel corpus di addestramento. Da qui, vengono selezionati migliaia di numeri attraverso cui la macchina costruisce un grande modello linguistico per token, producendo, successivamente, il linguaggio, prendendo un token dopo l'altro in maniera predittiva, statistico-probabilistica. Questo comporta l'ottenimento di risultati diversi di volta in volta, mantenendo sempre un output linguistico sintatticamente e semanticamente ottimo. Il fatto che nel titolo l'autore scelga di utilizzare una congiunzione coordinante e non una disgiuntiva ci fornisce già l'idea che questi due sistemi, per quanto diversi siano, coesistono anche per la correlazione tra di essi e, per questo, non viene necessariamente richiesta la sussistenza dell'uno senza l'altro.

Il primo sistema analizzato da Roncaglia è l'enciclopedia. Essa è uno degli strumenti che più è stato colpito dall'avvento di internet e della suddetta terza rivoluzione. Basti pensare che dalle semplici enciclopedie cartacee si è giunti alle prime forme di enciclopedie digitali, dopo un breve intermezzo di opere miste. La meglio riuscita e, quindi, la più diffusa nel mondo è Wikipedia, nata nel 2001 dall'esempio di Encarta. È stato ottenuto questo esito soprattutto per alcune caratteristiche intrinseche delle enciclopedie digitali, in particolare il costo di gran lunga inferiore rispetto a quelle

cartacee, che poteva ammontare ad alcune centinaia di dollari, e la possibilità di aggiornarle continuamente e di collegare le singole voci ad altrettante con un semplice click. Ciò che, inoltre, distingue Wikipedia dalle altre enciclopedie digitali è la modalità di costruzione delle note: essa può essere modificata non solo da esperti del settore, ma da qualsiasi utente, il che la rende una sorta di enciclopedia "democratica". Se da un lato questo avvicina gli utenti al mondo enciclopedico, dall'altra fa sorgere spontaneamente il dubbio sulla sua affidabilità. A questo sospetto, però, Roncaglia risponde in un'ottica positiva. L'attendibilità di Wikipedia risiede nel fatto che, nonostante la presenza dei cosiddetti "vandali" o "troll", il cui scopo è proprio quello di diffondere disinformazione, il sistema consente di individuare immediatamente le ultime modifiche e di riportare la voce alla versione precedente o di modificare il testo. Inoltre, in caso di necessità, è possibile bloccare sia l'indirizzo di rete del computer usato per scrivere il testo errato, sia la voce stessa. In aggiunta a ciò, maggiore è il numero di collaboratori, maggiore è anche la possibilità di avere tra essi esperti competenti, il che si sta verificando sempre di più. Un aspetto interessante da valutare è che esistono cinque pilastri su cui si basa Wikipedia e a cui i suoi collaboratori devono aderire: il fatto che sia un'enciclopedia; che deve essere scritta in maniera neutrale; che è un contenuto libero e chiunque può modificarlo, utilizzarlo, distribuirlo; che ha un codice di condotta, il che significa che è necessario ci sia un rapporto di collaborazione e non conflittuale; che non ha regole rigide, se non questi cinque punti.

Per quanto riguarda il secondo modello, "l'oracolo", esso comprende in particolare le intelligenze artificiali generative. Definite "pappagalli stocastici" dalla linguista Emily Bender, Roncaglia preferisce definirle "oracoli" poiché questi sistemi non copiano le informazioni che riescono a trovare in rete, come vorrebbe la suggestiva immagine dei pappagalli, ma li rielaborano, creando una composizione completamente originale e unica. Tuttavia, sia Bender che Roncaglia concordano sui problemi che questi strumenti intelligenti comportano: i costi sia economici che ambientali per la loro realizzazione e le allucinazioni e i pregiudizi prodotti. Un dato di fatto, che, nonostante la sua gravità, non è sufficiente per pensare di poter tornare indietro e di invertire la direzione di marcia di questa rivoluzione digitale.

Per chiarezza, ci soffermiamo sul problema delle allucinazioni e dei pregiudizi. Premettendo che le intelligenze artificiali necessitano di un addestramento che è in parte autonomo, in parte supervisionato, le risposte, essendo probabilistiche, a volte risultano errate. Ad esempio, a una domanda specifica su un fenomeno non esistente in una determinata epoca storica, l'IA si inventa una risposta congruente, ma non realistica: si tratta della cosiddetta "allucinazione". Posto questo problema, dettato anche da tentativi di inganno e di ricerca di punti deboli del sistema da parte degli utenti, nelle ultime versioni si è cercato di limitare tali errori, costruendo delle risposte pur sempre probabilistiche, ma più neutre, del genere "mi dispiace, ma c'è un errore nella tua richiesta".

Ciò dimostra, nelle nuove versioni, che con addestramenti sempre più specifici le allucinazioni dettate da errori fattuali si limiteranno sempre di più. Invece, i bias, ovvero i pregiudizi o distorsioni sistematiche nelle risposte, sono molto presenti. Basti pensare che alla domanda sull'identità di un fisico, la risposta prevedrà che questo sia sicuramente un uomo, mentre se si parla di un infermiere sarà donna. In risposta a questa questione, il Parlamento europeo ha approvato nel giugno 2023 l'AI Act, la proposta di regolamento dell'IA. Tra i punti più importanti c'è anche quello che richiede l'attenzione «all'attenuazione di possibili distorsioni nei set di dati, che potrebbero comportare rischi per i diritti fondamentali o risultare discriminatori per le persone interessate dal sistema di IA ad alto rischio [...] I risultati forniti da sistemi di IA sono influenzati da tali distorsioni intrinseche, che sono destinati ad aumentare gradualmente e quindi a perpetrare e amplificare le discriminazioni esistenti, in particolare nei confronti delle persone che appartengono a determinate minoranze vulnerabili o etniche o comunità razziali»[1].

Ulteriori problematiche degne di nota coinvolgono: «le politiche (o l'assenza di politiche) di gestione

dei dati personali forniti dagli utenti attraverso i loro prompt e il tema della possibilità di chiedere l'eliminazione di alcuni di tali dati da parte dei diretti interessati; la questione - presente nel secondo fra i passi della proposta di regolamento europeo citati sopra - del copyright di contenuti generati sulla base di un addestramento su dati creati da persone che non sono necessariamente consapevoli dell'inclusione all'interno del corpus di testi o immagini da loro creati, e che non partecipano né alla definizione delle relative politiche né agli utili derivanti dal loro sfruttamento economico; il rischio che la concorrenza sregolata fra società che producono sistemi di IA generativa si traduca in una eccessiva accelerazione nel loro sviluppo senza il tempo necessario alla riflessione e all'elaborazione di adeguate garanzie etiche, così come di valutazioni del loro impatto non solo culturale ma anche politico, sociale, economico e occupazionale» (pp. 104-105).

Entrambi i modelli sviluppati si inseriscono a pieno titolo, come già anticipato, nella rivoluzione digitale. Nonostante essa abbia consentito nel tempo una serie di miglioramenti della qualità della vita umana - basti pensare al periodo pandemico nel 2020, quando gli unici strumenti in nostro possesso per rimanere in contatto con l'esterno erano i social network e le applicazioni per videoconferenze o alla tendenza ormai irrinunciabile a cercare informazioni in modo semplice e veloce, in qualsiasi momento, su internet -, tuttavia vi sono anche degli effetti che possono essere pericolosi sia a livello sociale e comportamentale, sia a livello di privacy (e questa tendenza aumenta esponenzialmente con l'avvento e la diffusione di strumenti di IA). Partendo dalla prima questione, si può facilmente osservare la tendenza dei social network a favorire contenuti a bassa complessità, il che causa chiaramente anche una diminuzione delle competenze di interpretazione di contenuti complessi, che, tuttavia, come sottolinea l'autore, sono indispensabili per l'umanità per poter comprendere il mondo circostante. Sarebbe chiaramente un dato allarmante, ma Roncaglia ritiene, in realtà, che tale tendenza alla granularità e alla semplificazione dei contenuti non sia una condizione permanente e caratterizzante dell'ecosistema digitale, ma una situazione contingente e attuale. Difatti, l'autore sostiene che già ora ci siano dei raggi di sole per un panorama più positivo in un non ben definito futuro. Le argomentazioni da lui portate si focalizzano sull'aumento dei secondi nei video dei principali social network, come TikTok (da 15 a 60 secondi a 3 minuti sino ai 10 nel 2022) o Instagram (da 15 a 60 secondi) e sulla diffusione della pratica di ascoltare podcast e audiolibri. Questo quadro, che rimane tuttavia ancora limitato e graduale, apre un'ulteriore questione degna di attenzione, ovvero il problema della capacità di comprensione della complessità. Essa non necessariamente è sinonimo di qualità e l'esposizione a questi contenuti rimane alta a partire dalla giovane età. Ciò comporta, chiaramente, la necessità di una formazione e, quindi, di un insegnamento alla lettura, alla comprensione e all'analisi di questi contenuti e il luogo più importante in cui tutto ciò può avvenire è senz'altro la scuola. Si rende necessario, dunque, organizzare un sistema scolastico che orienti le sue politiche a questo fine. Sicuramente è una prospettiva auspicabile e fondamentale, nella speranza che in un Paese in cui i tagli all'istruzione diventano sempre più pesanti e allarmanti si voglia realmente attuare un programma di ulteriore apprendimento. L'unico passo fatto finora è il Piano Nazionale Scuola Digitale, che fa parte del progetto "Buona Scuola" a cui ha partecipato nel 2021 lo stesso Roncaglia, ma ad oggi il tutto parrebbe insufficiente.

Un aspetto sicuramente positivo che fa da contraltare al discorso esposto finora è il nuovo impatto che le IA hanno avuto sulla memoria, che senza alcun dubbio costituisce una caratteristica fondamentale per la costruzione della nostra identità. Si è notato che esiste una correlazione molto stretta tra la memoria interna, legata al nostro cervello biologico e la memoria esterna che, invece, è dettata da strumenti, perlopiù tecnologici. A questo proposito, Andy Clark e David Chalmers nel 1998 hanno sviluppato una teoria, detta "della mente estesa"[2] che sottolinea come «la nostra mente utilizzi per i suoi processi cognitivi non solo il cervello biologico, ma molte sue estensioni esterne, naturali e sempre più spesso artificiali o culturali» (p. 126). I memofilm, solo per fare un

esempio, si sono rilevati molto benefici per le persone che soffrono di amnesia o altri disturbi della memoria.

Da questo quadro, in conclusione, traspare il fatto che la terza grande rivoluzione digitale è - e, per una naturale tendenza, sarà - un fenomeno molto importante per la vita dell'uomo e per l'ambiente circostante, ma anche per l'accesso alla conoscenza e alla divulgazione del sapere, grazie alla sua capacità di offrire significativi ausili, ma il dipinto roseo potrà essere mantenuto solo se non verranno sottovalutati i rischi che sin da ora emergono e se ci si muoverà senza indugi con le istituzioni verso una regolamentazione e un controllo concreti, evitando che la strada del progresso diventi, invece, un percorso irreversibile di regressione collettiva.

[1] Emendamenti del Parlamento Europeo, approvati il 14 giugno 2023.

[2] Si veda l'articolo di Andy Clark e David Chalmers, *The Extended Mind*, in «Analysis» 58,1, gennaio 1998.