

Intelligenza Artificiale: Umana troppo umana.

Quando si affronta un'innovazione tecnica che colpisce l'immaginazione per le sue ricadute rivoluzionarie nella vita individuale e collettiva, è facile cadere in considerazioni fantasiose dettate da speranze o timori, trascurando un'analisi razionale e tecnica dell'innovazione stessa.

Per evitare tali errori è necessario cercare di spiegare nel modo più semplice e intuitivo possibile i principi che stanno alla base della tecnologia in questione.

L'Intelligenza Artificiale (IA), insieme all'emergenza climatica, è in questo momento l'argomento che domina i dibattiti sul nostro futuro. Non a caso, da molti è considerata "una singolarità tecnologica".

Rispetto ad altre innovazioni tecnico-scientifiche del XX secolo che hanno cambiato in modo radicale il nostro modo di vivere, ha la particolarità di promettere mutamenti radicali non solo nella sfera del lavoro manuale e in quella del lavoro intellettuale strettamente ripetitivo – situazione alla quale siamo abbastanza abituati con la diffusione capillare dei computer – ma promette/minaccia di sostituire l'uomo anche nelle attività di alto livello, di ricerca e creative.

Sulla IA si fronteggiano opinioni contrapposte: alcune la vedono come uno strumento salvifico, destinato a realizzare in terra un paradiso "transumano" nel quale saremo sollevati da ogni incombenza fisica e mentale; altre profetizzano un futuro oscuro, nel quale la IA, dopo aver superato le capacità intellettive della specie umana, la ridurrà in schiavitù o addirittura la eliminerà del tutto.

La radicale frattura nelle aspettative sulla IA non vede contrapporsi gli esperti all'uomo della strada, ma attraversa l'intera comunità scientifica ai massimi livelli, compresi i pionieri della IA e le grandi aziende d'informatica che hanno puntato sul suo sviluppo.

Ridurre tali contrapposizioni a fantasie su oscuri complotti o a smanie di protagonismo è semplicistico e non coglie l'aspetto di fondo del fenomeno.

Anche in altre situazioni, come l'uso civile del nucleare o le origini delle pandemie, è capitato di assistere a scontri tra esperti.

Ma nel caso della IA vi è qualcosa di particolare, in quanto non si tratta solo dei problemi della perdita di posti di lavoro o della difesa della privacy, ma della capacità di sostituire l'attività umana in settori fino ad ora ritenuti di alto livello.

Quello che stupisce non è la capacità di queste applicazioni di gestire grandi quantità di informazioni ma di essere più brave di molti umani nel riassumere o generare un testo, generare immagini e filmati, produrre del codice nei vari linguaggi di programmazione sulla base di descrizioni del problema espresse in lingua naturale.

È molto singolare che gli avvertimenti più allarmanti vengano da studiosi che sono stati tra i fondatori della IA e da responsabili delle società impegnate nel suo sviluppo, di solito le cose non vanno in questo modo.

Non è da escludere che dietro a queste posizioni si nasconda in ultima analisi l'orgoglio misto a paura di aver creato qualcosa che insidia la superiorità intellettuale della specie umana.

Emblematico di questo atteggiamento confuso nei confronti della IA è l'uso generalizzato del termine "algoritmo" in tutte le pubblicazioni, sia specialistiche che divulgative sulla IA.

La locuzione "L'algoritmo ha deciso ...", quando non è solo un modo per sottrarsi alle responsabilità di scelte scomode, suona come evocativo di un'autorità superiore, infallibile come le leggi della matematica.

Ma l'uso di questo termine per IA è un inganno semantico.

Fin dalle elementari l'algoritmo viene definito come una sequenza ordinata e finita di istruzioni che eseguono un processo in modo deterministico.

Detto alla buona, una divisione fra due numeri dati, se correttamente eseguita, deve dare sempre lo stesso risultato.

Nel caso della IA la situazione è diversa, i modelli linguistici alla base delle applicazioni di IA sono addestrati con miliardi di parametri e il processo che genera una risposta si basa sulla statistica e la probabilità.

E anche se i processi di addestramento si basano su degli algoritmi, il processo di interpretazione delle domande e generazione delle risposte si basa su una semantica delle parole che non è affatto riconducibile a rigorose definizioni ma è qualcosa che si autodefinisce all'interno di reti neurali artificiali attraverso meccanismi di tipo statistico.

I fondamenti teorici della IA e delle reti neurali risalgono agli anni 50, ma sono stati abbandonati per seguire altre vie perché ritenuti inutilizzabili.

Solo a partire dalla fine degli anni 90 l'incremento della capacità di calcolo e la disponibilità di enormi quantità di dati hanno permesso un salto qualitativo nell'utilizzo delle reti neurali per arrivare infine ai risultati eclatanti dei nostri giorni.

Nelle applicazioni nel campo del linguaggio, che è il settore che più colpisce l'immaginazione, il primo passo è quello di trasformare le parole e le frasi in liste di numeri; il presupposto teorico è piuttosto semplice.

Si fonda sull'ipotesi di distribuzione la quale afferma che le parole che appaiono in contesti simili tendono ad avere significati simili.

In altri termini, il significato di una parola può essere compreso osservando le parole che tendono ad apparire vicino a essa in vari contesti.

Questo principio è stato enunciato da John Rupert Firth, un linguista britannico, che ha affermato: "You shall know a word by the company it keeps" (Conoscerai una parola dalla compagnia che frequenta)

Quindi analizzando un testo per ogni parola si assegnano dei numeri a queste liste sulla base dei termini che precedono e seguono ogni parola.

Questo processo si chiama embedding (incorporamento); senza scendere in particolari tecnicistici va rilevato che la cosa importante è che si passa da una analisi del testo di tipo

lessicale e morfologica ad una di tipo semantico.

È paradossale ma indicativo di un atteggiamento orfano del rassicurante modello meccanicistico, che si chiami “algoritmo” qualcosa che si basa su processi statistico e può essere messo a punto solo per tentativi e le cui risposte non sono del tutto prevedibili.

Un fatto è certo, la IA funziona, produce testi, poesie, immagini, risponde con competenza a domande complesse, fa diagnosi mediche, guida le auto; e siamo solo agli inizi.

Qualche volta sbaglia, ma questo rientra nelle attese; un'approssimazione intorno al 90% nell'esattezza dei risultati è considerato un successo; una del 100% una prova che si è sbagliato qualcosa nell'addestramento o nei test, proprio perché non è un algoritmo.

Supponiamo che per una stranissima evoluzione della scienza e della tecnica si sia pervenuti a sviluppare una forma di IA prima che diventasse lecita e diffusa la teoria copernicana.

La nostra IA sarebbe stata addestrata sui testi disponibili in tutte le biblioteche valutandone l'affidabilità sulla base dell'autorevolezza degli autori.

Siccome anche allora esisteva il politicamente corretto, durante la sua messa a punto le avrebbero insegnato a non dare risposte con risvolti immorali o blasfemi.

Al seminarista inquieto che avesse fatto domande sulle stagioni o sull'alternanza del giorno e la notte, la IA avrebbe “correttamente” risposto secondo l'affidabile teoria Tolomaica.

La distanza fra la conoscenza degli strumenti ed il loro uso è una caratteristica della modernità.

Con la IA tale distanza è aumentata a dismisura, e visto la sua pervasività in tutti i campi ad ogni livello, questa situazione è oggettivamente preoccupante.

Non si tratta di pretendere che tutti diventino degli esperti informatici ma che vi sia un minimo di consapevolezza sulla natura della IA.

Comprendere la sua natura probabilistica è fondamentale, siamo abituati a una fiducia assoluta nell'affidabilità di calcolo dei computer, trasferire la stessa fiducia nelle applicazioni della IA può essere molto pericoloso.

È fin troppo simile a noi, anche nello sbagliare.