

Luci e ombre dell'intelligenza artificiale

di Silvano Marioni

Il filosofo e matematico Alfred North Whitehead ha scritto che i maggiori progressi della civiltà sono i processi che distruggono le società in cui si verificano. Le nostre generazioni sono testimoni di alcune di queste rivoluzioni che hanno portato cambiamenti fondamentali. Ad esempio, internet e gli smartphone hanno condizionato molti aspetti della nostra vita quotidiana, modificando le nostre abitudini in modo importante, sia nel modo di lavorare che nelle dinamiche sociali. Quando la tecnologia diventa semplice da usare è generalmente destinata ad avere successo. Sarà così anche per l'intelligenza artificiale?

Nata come disciplina accademica nella metà degli anni '50, l'intelligenza artificiale (IA) è stata caratterizzata da un alternarsi di periodi di entusiasmo e di delusione fino quando, alla fine del 2022, il fenomeno ChatGPT ha riaccessato in maniera definitiva l'interesse dell'opinione pubblica e degli investitori.

Cos'è ChatGPT?

È un programma informatico in grado di generare dei testi simili a quelli umani, che a fronte di domande anche di una certa complessità, è in grado di fornire le risposte adeguate in varie lingue, italiano compreso.

ChatGPT è specializzato nella comprensione ed elaborazione del linguaggio, come ad esempio in tutti quei casi in cui si vuole ricercare informazioni, generare o correggere testi, riassumere documenti, fare traduzioni, creare testi narrativi, racconti, programmi informatici e molto altro ancora. Può ad esempio suggerirci un testo originale per un biglietto di auguri o riscrivere in modo più chiaro una frase che non ci soddisfa, ma anche proporci una ricetta di una torta per una persona celiaca o pianificare per noi i periodi di semina nel nostro orto. ChatGPT ha la capacità di generare testi con una varietà di toni diversi che vanno dall'informale al persuasivo, dall'umoristico al professionale. Ha anche la capacità di adattarsi a diversi stili di scrittura, producendo testi nello stile di questo o quello scrittore.

Esistono altri sistemi di IA simili a ChatGPT ma sono meno noti, avendo avuto finora minore risalto sui media.

Le basi dell'IA

Capire il funzionamento di questi sorprendenti sistemi non è immediato perché si basano su tecnologie complesse, modelli matematici avanzati, e raffinate tecniche di apprendimento automatico. Se volessimo semplificare il loro funzionamento potremmo dire che si basano su due elementi principali: le tecniche di IA generativa e i modelli linguistici di grandi dimensioni.

Le tecniche di IA generativa permettono di comprendere le domande attraverso l'analisi del linguaggio per iniziare a dare come risposta una parola relativa al contesto della domanda. Poi prevedono la parola successiva, e quella successiva ancora e continuano, parola dopo parola, fino a quando è stata generata la risposta completa. In parole povere sono sistemi in grado di prevedere le parole successive in modo quasi statistico e relativo al contesto.

I modelli linguistici di grandi dimensioni sono invece dei sistemi dotati di capacità di apprendimento che vengono "addestrati" con enormi quantità di dati testuali provenienti da fonti diverse. Nel caso di ChatGPT questo processo viene ripetuto milioni di volte, permettendo al modello di apprendere in modo autonomo molti schemi e relazioni tra parole e frasi nella lingua in cui viene istruito. Le stime indicano che solo per l'addestramento di ChatGPT siano state utilizzate centinaia di miliardi di parole ricavate da libri, articoli di giornale, pagine internet e conversazioni con gli utenti, con un volume di testo equivalente a quelle di 45 milioni di libri. In alcuni casi è stato necessario fornire un aiuto umano al sistema di addestramento, suggerendo risposte per migliorare l'allineamento delle conoscenze o per eliminare argomenti scabrosi come le descrizioni di violenza, terrorismo, abusi sessuali, omicidi e altre attività criminali.

I programmi di IA avanzata richiedono importanti attività di ricerca e supercomputer capaci di elaborare miliardi di parametri e sono al momento alla portata solo delle aziende con grandi risorse tecniche e finanziarie. OpenAI, la società che ha creato ChatGPT, non ha mai reso noto i costi di sviluppo del sistema, ma si sa che dispone di numerosi finanziatori miliardari, tra cui anche Microsoft.

Risposte accurate sì ma...

Come fa ChatGPT a essere così abile a conversare e a rispondere alle nostre domande? Per ottenere risultati accurati, le domande devono essere poste in modo preciso, fornendo dettagli sul contesto e l'argomento. Più informazioni vengono date, più il sistema può focalizzarsi e dare risposte corrette. Se la risposta iniziale non è abbastanza chiara, si possono fornire nuovi dettagli per guidarlo ad essere più preciso.

Questi sistemi IA hanno delle conoscenze linguistiche molto avanzate, ma possono talvolta fornire informazioni inesatte. Se chiedo informazioni su un inesistente Convento del Monte Boglia a Lugano ricevo una descrizione linguisticamente corretta e dettagliata di un fantomatico convento francescano. Tuttavia, una ricerca online conferma che non esiste alcun convento con quel nome a Lugano. Il sistema cerca di darci una risposta puramente linguistica anche nei casi in cui non dispone di informazioni o quando noi ci aspettiamo delle risposte in situazioni basate sul caso o la probabilità. Questi casi di errori dei sistemi di IA, noti con il termine di "allucinazioni", sono lacune alle quali gli informatici



stanno già lavorando per migliorare i modelli e ridurre quindi al minimo le inesattezze. Del resto la correttezza dei contenuti creati da questi programmi non è garantita e, proprio per questo, i fornitori dei sistemi di IA raccomandano gli utenti di non considerare affidabili le risposte relative a consigli medici, legali o finanziari.

I sistemi di IA sono invece molto efficaci nel produrre e rielaborare testi, ad esempio per generare documenti, riassumerli o revisionarli. Si possono porre domande di carattere generale, come ad esempio chiedere quali fiabe hanno una morale simile a "La cicala e la formica", e ottenere delle indicazioni utili.

La responsabilità finale di valutare l'accuratezza e l'affidabilità delle risposte resta comunque sempre a noi, alla nostra capacità di valutazione e al nostro buon senso.

Dubbi e timori

L'uso dei sistemi di IA solleva anche alcune preoccupazioni etiche. Ad esempio, non è possibile spiegare come i sistemi arrivino alle loro conclusioni e questo crea incertezze sulla responsabilità delle società che li sviluppano e su quali siano i loro reali obiettivi. La capacità di generare testi realistici e credibili può inoltre creare disinformazione in modo più subdolo e pericoloso di quanto abbiano fatto finora i social network, minando la fiducia del pubblico, influenzando le opinioni politiche e favorendo nuove forme di truffe.

I sistemi di IA memorizzano le conversazioni con gli utenti per ampliare le proprie conoscenze e migliorare le prestazioni. Ci sono quindi dei rischi per la sicurezza e la privacy se sono utilizzati per elaborare informazioni riservate o personali. Chi volesse fare la revisione di un documento aziendale riservato, finirebbe per condividere i suoi dati sensibili che potrebbero diventare accessibili anche a terze parti.

Da ultimo c'è il tema dei diritti d'autore. Se ChatGPT riesce a scrivere una poesia con lo stile di Giuseppe Ungaretti vuol dire che, in fase di addestramento, deve aver "letto" e memorizzato tutte le sue opere protette dal diritto di autore. Stanno quindi iniziando i primi dibattiti sulle modalità indiscriminate di raccolta delle informazioni da parte delle società di IA e sono state avviate le prime azioni legali sulle violazioni dei diritti di autore. Per concludere, il rischio principale di fronte a questa tecnologia emergente è quello di riporre in essa delle aspettative eccessive che ci fanno confondere le capacità dei sistemi di IA con le emozioni e pensieri tipicamente umani. Il problema più importante non è se questi sistemi possano diventare più intelligenti di noi, ma come noi li utilizzeremo e fino a che punto, per comodità o pigrizia, lasceremo che prendano decisioni importanti al nostro posto.