L'intelligenza artificiale può mentirci?

Martina Barni

Studentessa di economia e scienze sociali presso l'Università Bocconi e membro dell'associazione studentesca Economic Society for Bocconi Students.

Federico Casotto

Studente di economia e finanza presso l'Università Bocconi e membro dell'associazione studentesca Economic Society for Bocconi Students.

Intervista a Tomaso Poggio

Uno dei fondatori delle neuroscienze computazionali e figura di spicco nel campo dell'intelligenza artificiale. Professore al Massachusetts Institute of Technology (MIT), è il direttore del Center for Brains, Minds and Machines (CBMM), un centro interdisciplinare che esplora le basi scientifiche dell'intelligenza.

L'intelligenza artificiale è una delle innovazioni tecnologiche che più rapidamente è passata dalle mani dei ricercatori a quelle degli utenti finali. Sebbene la platea di utilizzatori di software come ChatGPT sia già ampia, spesso fatichiamo a comprendere cosa sia effettivamente l'IA. E cosa sia in grado di fare. Dopo cinquant'anni in cui informatica voleva dire programmare i computer, cioè impartire loro istruzioni, negli ultimi venti abbiamo assistito a un radicale cambio di passo: le macchine hanno cominciato ad apprendere da esempi e da tempo sono in grado di battere i migliori giocatori umani di scacchi. I progressi sono tangibili anche nelle applicazioni più pratiche, come nella guida autonoma, seppur ancora con delle limitazioni. Su tutto questo si inserisce il grande tema della regolamentazione, ad esempio sul fronte della privacy, fra crescenti preoccupazioni dell'opinione pubblica, oltre che sulla questione della trasparenza degli algoritmi. Infine, ci sono ancora tanti interrogativi aperti sui limiti di queste macchine. Sono coscienti? Possono ingannarci? Sono domande che stimolano riflessioni più ampie che abbracciano il tema della conoscenza e intelligenza umana, in un vasto discorso che intreccia filosofia, neuroscienza computazionale e informatica. La ricerca interdisciplinare di Tomaso Poggio spazia dal cervello umano ai computer ed è guidata dall'idea che conoscere le dinamiche dell'apprendimento sia fondamentale per comprendere sia l'intelligenza biologica sia quella artificiale. L'IA sta già trasformando il nostro mondo in modi che solo pochi anni fa sembravano fantascienza. Approfondire le sfide e le opportunità future dell'intelligenza artificiale con uno dei massimi conoscitori delle potenzialità e dell'avanzamento degli studi in questo settore offre una visione stimolante e inedita sull'oggi e sul domani di questa tecnologia.

Professor Poggio, oggi ovunque si parla di intelligenza artificiale, a volte anche a sproposito. A uno dei padri dell'IA vorremmo chiedere di spiegarci cosa sia.

L'intelligenza artificiale rientra tra quelle che Marvin Minsky ha definito "*suitcase words*", parole bagaglio, che non hanno senso in sé ma contengono molti significati. Definirei l'intelligenza artificiale come il tentativo di riprodurre nelle macchine l'intelligenza umana. Poi, negli ultimi anni, l'IA si è sostanziata nel *machine learning*, ovvero il tentativo di far imparare le macchine.

Quella dell'intelligenza artificiale è una vera e propria rivoluzione scientifica. C'è un momento particolare in cui si è accorto che stava nascendo una nuova scienza?

Nei primi cinquant'anni della storia dei computer, l'informatica si è occupata di programmarli. Negli anni Duemila c'è stato un cambio di paradigma, una vera e propria rivoluzione scientifica: dal programmare le macchine siamo passati all'insegnare alle macchine, che imparano da esempi (il cosiddetto *deep learning*).

Quand'è che le macchine hanno iniziato a competere con gli umani?

Vi faccio un esempio. DeepMind, una società fondata nel 2011 da un ricercatore con cui ho collaborato, è nata con l'idea di costruire un sistema di intelligenza artificiale basato sull'apprendimento, che sapesse giocare a qualunque gioco. Nel 2015, è stato sviluppato un sistema chiamato AlphaGo, che in Corea del Sud ha battuto il campione mondiale ufficioso di Go, un gioco molto più complesso degli scacchi e con una lunga tradizione in Asia. Durante una di queste partite, AlphaGo ha eseguito una mossa incredibile, di quelle che vengono definite "God's moves" (mosse divine), così sorprendente e creativa da essere stata studiata per anni. Questo è un grande successo perché AlphaGo è in grado di imparare a giocare a scacchi, Go e altri giochi autonomamente, semplicemente giocando contro se stesso. In sei ore può diventare più bravo del campione mondiale di scacchi e in ventiquattro ore può diventare migliore di quello di Go. Tutto ciò dimostra che nel mondo virtuale dei giochi l'intelligenza artificiale è già superiore a quella umana.

Oggi le intelligenze artificiali sono in grado di risolvere problemi reali. Come è avvenuto questo cambio di passo?

Una delle principali sfide è indubbiamente l'applicazione dell'IA nel mondo reale. Mobileye, fondata da un mio ex collaboratore, Amnon Shashua, è la società che ha fornito i primi sistemi di guida automatica per Tesla. Recentemente, ha dovuto interrompere la fornitura perché Elon Musk non spiegava adeguatamente ai suoi clienti i limiti di queste macchine, tanto che in Texas una persona è morta. Oltre alla sfida di creare sistemi più che sicuri se applicati nel mondo reale, queste macchine devono essere istruite per gestire situazioni fuori dall'ordinario. Mobileye ha sviluppato un sistema che funziona molto bene, anche nel traffico disordinato di Gerusalemme. Tuttavia, questi sistemi dipendono da mappe precise e potrebbero non funzionare in aree non mappate. Inoltre, potrebbero non essere in grado di gestire situazioni particolari, come incidenti o ordini emanati dalle autorità.

Lei lavora negli Stati Uniti. Secondo lei, esiste un gap nella ricerca e nello sviluppo riguardo l'intelligenza artificiale tra Europa e Usa?

Sì, esiste. Direi che attualmente le due superpotenze nell'intelligenza artificiale sono gli Stati Uniti e la Cina. Il *know-how* esiste anche in altre parti del mondo, come in Italia e in Europa, e quello che manca non è la qualità dell'insegnamento nelle università, che è comparabile a quella americana. Il problema è che in Europa mancano società con risorse e personale sufficienti per avere un vero impatto. Ad esempio, DeepMind, la società cui ho accennato prima, è stata fondata più di dieci anni fa a Londra. Recentemente è stata acquisita da Google e ora a Londra lavorano mille persone, mentre tra San Francisco e Mountain View ce ne sono ben 2500.

E come devono cambiare le società europee per chiudere il divario?

Per avere un impatto significativo nel campo dell'intelligenza artificiale è necessaria una concentrazione di risorse. Non basta avere un singolo ricercatore brillante. Le idee interessanti possono ancora venire da singoli ricercatori, ma per sviluppare prodotti sono necessarie società con grandi risorse umane e computazionali, così come investimenti finanziari significativi. Insomma, i fondi raccolti da startup come OpenAI sono impressionanti. Parliamo di centinaia di miliardi di dollari. In Europa, ci sono alcune iniziative e qualche startup promettente, in particolare in Francia, ma l'Italia è indietro. So che alcuni fondi stanziati per l'IA in Italia si sono dispersi in troppi progetti e questo è un errore perché serve concentrare risorse per avere un impatto significativo.

Un altro aspetto rilevante, che sembra preoccupare l'opinione pubblica, sono i rischi dell'IA. Tra questi c'è la creazione di contenuti falsi ma estremamente realistici, i cosiddetti "deep fake". Come si può ridurre il rischio di disinformazione?

Certo, ci sono rischi evidenti. Non è solo un problema di *deep fake*, ma anche di proprietà intellettuale, perché si dovrebbe sempre dichiarare esplicitamente da dove proviene il contenuto di un materiale diffuso. Dal punto di vista pratico, c'è una difficoltà più grande nel caso dell'IA, comune anche ad altre tecnologie. Mentre tecnologie come l'energia nucleare e le bombe atomiche possono essere facilmente limitate, perché sono molto onerose le infrastrutture richieste per implementarle, l'IA è accessibile a tutti. Possiamo proibire certe pratiche in Europa, ma qualcuno le adotterà in Corea, in Cina o in Africa. Ormai, il genio è fuori dalla bottiglia e non si può fermare. Per questo non dobbiamo scordarci che i rischi esistono, ma ci sono anche grandi benefici potenziali e la sfida politica è forse più complessa dei problemi tecnologici e scientifici.

Un'altra preoccupazione riguarda la privacy degli utenti. I large language models, come per esempio il motore alla base di ChatGPT, vengono allenati utilizzando grandi quantità di dati. Concretamente, come possiamo garantire che i dati personali degli utenti del web rimangano privati?

Concordo con l'attenzione che dell'Unione europea riserva alla privacy. Si tratta chiaramente di un pericolo reale, che non riguarda solo l'IA e modelli come ChatGPT, ma tutte le grandi aziende, come Google, che raccolgono enormi quantità di informazioni su ciascuno di noi senza averci mai veramente chiesto il permesso in modo adeguato.
Dovremmo avere il diritto alla privacy e il diritto ai nostri dati. È un diritto che diventerà sempre più importante in futuro. Una delle richieste della Comunità europea è che, se una persona ritiene che i propri dati siano stati utilizzati senza il suo consenso e desidera che non vengano più utilizzati, dovrebbe essere possibile cancellarli. Per i sistemi di addestramento attuali, come quelli utilizzati per GPT, è un grande problema perché non è possibile rimuovere i dati senza dover ricominciare l'intero processo di addestramento, che dura mesi. Tuttavia, potrebbero nascere nuove soluzioni tecnologiche in futuro e sarebbe un progresso sostanziale per il rispetto dei diritti degli utenti.

Esiste anche il problema della mancanza di trasparenza negli algoritmi, che impedirebbe di individuare e correggere risultati sbagliati o eticamente ingiusti. Pensa che gli standard attuali siano sufficienti?

È essenziale migliorarli. La mancanza di trasparenza negli algoritmi è un problema significativo per l'intelligenza artificiale e per molte altre tecnologie avanzate. La pubblicazione degli algoritmi sarebbe importante non solo da un punto di vista di trasparenza verso gli utenti, ma sarebbe anche utile per consentire un avanzamento nella ricerca.
A differenza di fasi precedenti, in questo momento sono poche le aziende che rendono pubblici i propri algoritmi. La stessa OpenAI, nonostante il nome "open", non è una di queste.

Nei sistemi di machine learning l'output è generato automaticamente dall'addestramento della macchina su grandi quantità di dati. Vede possibilità di coscienza in questo processo?

È un tema rilevante. Innanzitutto, non è chiaro cosa intendiamo con coscienza. Una forma di coscienza è l'autocoscienza, ovvero la consapevolezza che un soggetto ha di se stesso. Ad esempio, un essere è considerato autocosciente quando è in grado di riconoscersi allo specchio (*mirror test*). Poche specie animali superano questo test. Chat GPT lo passerebbe? Gliel'ho chiesto e ha dichiarato di poterlo passare, ma solo perché sa di cosa si tratta. Sarebbe interessante rimuovere ogni riferimento al *mirror test* dall'insieme di dati che alimentano Chat GPT per vedere se passerebbe davvero il test.

Esistono altri modi per testare la coscienza di ChatGPT?

Mi sono chiesto, ad esempio, se ChatGPT mente. ChatGPT è in grado di "ingannarci" se gli chiediamo esplicitamente di farlo e, ad esempio, di rispondere fingendo di essere una determinata persona. Ma se ci mentisse o ci nascondesse delle risposte intenzionalmente, un po' come accade nel film "Ex machina" (che vi consiglio), allora sarebbe effettivamente cosciente e noi non abbiamo ancora un modo per scoprirlo.

L'intelligenza artificiale è nata dallo studio dell'intelligenza umana. La somiglianza tra reti neurali umane e artificiali è l'esempio più intuitivo. Pensa che ulteriori progressi necessitino di altri prestiti dalla biologia o l'evoluzione si basa su altro?

Fino a dieci anni fa ero convinto che per fare scoperte sulle macchine avremmo dovuto passare da avanzamenti nello studio della mente

umana. La neuroscienza ha guidato i primi passi nella creazione di macchine intelligenti. Per intenderci, le *neural networks* e algoritmi come *deep learning* nascono dalla neuroscienza. Negli ultimi sette anni, invece, il progresso nell'IA è avvenuto senza prendere spunto dalla neuroscienza, perché l'ingegneria ha potenziato gli algoritmi esistenti senza più ispirarsi al funzionamento del cervello umano. Mi riferisco, ad esempio, ai *large language models* che hanno portato allo sviluppo di ChatGPT.
Non escludo che nei prossimi anni potremmo avere bisogno di tornare alla neuroscienza per avere nuove ispirazioni.

È possibile il contrario? Ovvero, è possibile che dallo studio della mente delle macchine si possano derivare scoperte sulla mente umana? È mai successo?

È la prima volta che ci sono intelligenze del nostro livello. Significa che ora possiamo fare studi comparativi tra l'intelligenza umana e le nuove intelligenze. Pensiamo, ad esempio, al Dna, che è la base genetica universale di quasi tutte le forme di vita conosciute. Allo stesso modo, se scoprissimo alcuni principi comuni tra forme di intelligenza, allora sì, potremmo fare scoperte sulla mente umana studiando l'intelligenza artificiale.

Sappiamo che gli esseri umani presentano dei "pregiudizi" (bias) cognitivi nel ragionamento. Lei pensa che l'intelligenza artificiale, per come è allenata, abbia ereditato questi difetti umani?

Certamente, le macchine hanno ereditato *bias* umani perché vengono addestrate con i prodotti della mente umana. È difficile dire se l'intelligenza artificiale abbia un limite posto da ciò che la mente umana ha conosciuto o se sarà capace di originalità nella matematica e nella produzione di nuove congetture.

Possiamo aspettarci che questo avvenga nei prossimi anni?

Sì, mi aspetto che in futuro avremo una macchina cosciente. Fino a qualche anno fa credevo ci sarebbero voluti almeno cinquant'anni. Ora penso che potrebbe succedere anche domani.

Quale sarà l'impatto sulla società?

Due secoli fa Alessandro Volta inventò la pila, che per la prima volta garantiva una fonte di elettricità continua; da quel momento in poi si sono rapidamente susseguite moltissime applicazioni. Anche il cambio di paradigma dal programmare i computer all'insegnare ai computer ha sicuramente aperto la strada a una serie innovazioni che dobbiamo aspettarci arrivino a ritmi velocissimi.
Uno dei grandi pericoli è che questa volta il cambiamento avvenga troppo rapidamente. Ci vuole del tempo per permettere alla società di adeguarsi. Quando siamo passati dai cavalli ai treni, molte persone che lavoravano con le carrozze hanno perso il lavoro, ma non è successo tutto in una volta. C'è stato un periodo di adattamento, che oggi sembra non trovare spazio.
Quale penso sia l'impatto sulla società? Non so dire se la Borsa sia troppo ottimista, ma certamente l'intelligenza artificiale porterà a una rivoluzione industriale.
È essenziale che la società tenga il passo e riesca ad adattarsi per trarre tutti i benefici che questa nuova scienza ha da offrirci.