DOPPIOZERO

Coscienza artificiale: l’ultima frontiera

Riccardo Manzotti
31 Gennaio 2024

Qualche giorno prima di Natale, sulla rivista *Nature* è stato pubblicato un articolo che fa capire quanto l’intelligenza artificiale stia ormai travolgendo molti assunti di fondo. Il titolo dell’articolo, scritto dalla giornalista scientifica Maria Lenharo, anche solo dieci anni fa sarebbe stato impensabile: «Coscienza artificiale: gli scienziati premono per avere risposte». Il fatto stesso che si prenda in considerazione la possibilità di una macchina cosciente è epocale perché implica la possibilità che la divisione culturalmente più profonda, quella tra noi e le macchine, sia messa in discussione. È straordinario vedere come una sfida tecnologica (l’intelligenza artificiale) e un problema scientifico (la coscienza) convergano nel pretendere una risposta univoca.

In sintesi, l’autrice sottolinea come le capacità cognitive dell’Intelligenza Artificiale, grazie ai famosi modelli generativi del linguaggio di grandi dimensioni (LLM) di cui ChatGPT è l’esempio di maggior successo, siano oggi talmente vicine a quelle umane da rendere plausibile la domanda se queste intelligenze artificiali siano coscienti.

Recentemente, i coordinatori della prestigiosa *Association for Mathematical Consciousness Science* hanno sottolineato come non esistano criteri oggettivi per riconoscere i sistemi dotati di coscienza. Già dall’Aprile 2023, Jonathan Mason (matematico a Oxford), Lenore Blum (studioso di intelligenza artificiale di fama mondiale) e Johannes Kleiner (matematico presso l’università di Monaco) hanno chiesto di prendere seriamente in considerazione la possibilità della creazione di macchine dotate di coscienza in una lettera aperta al Consiglio sull'Intelligenza Artificiale delle Nazioni Unite.

È un tema ricorrente. Nel 2022 un ingegnere della Google, Blake Lemoine, aveva dichiarato che l’intelligenza artificiale *Lambda* (oggi ampiamente superata da ChatGPT) fosse autocosciente. Non era stato preso sul serio, ma in meno di due anni una dichiarazione azzardata è diventata una domanda rispettabile. Oggi più di uno studioso prende seriamente in considerazione la possibilità della coscienza artificiale. C’è chi propone di sospendere il progresso tecnologico in attesa di capirne di più e c’è chi sottolinea i problemi etici, nonché i rischi, posti anche solo dalla possibilità di realizzare una coscienza artificiale. Oggi che le macchine stanno raggiungendo gli esseri umani sia per quanto riguarda il linguaggio che per quanto riguarda la capacità di muovere corpi e di interagire con il mondo, tra macchine e persone l’ultima differenza è rappresentata proprio dalla coscienza.

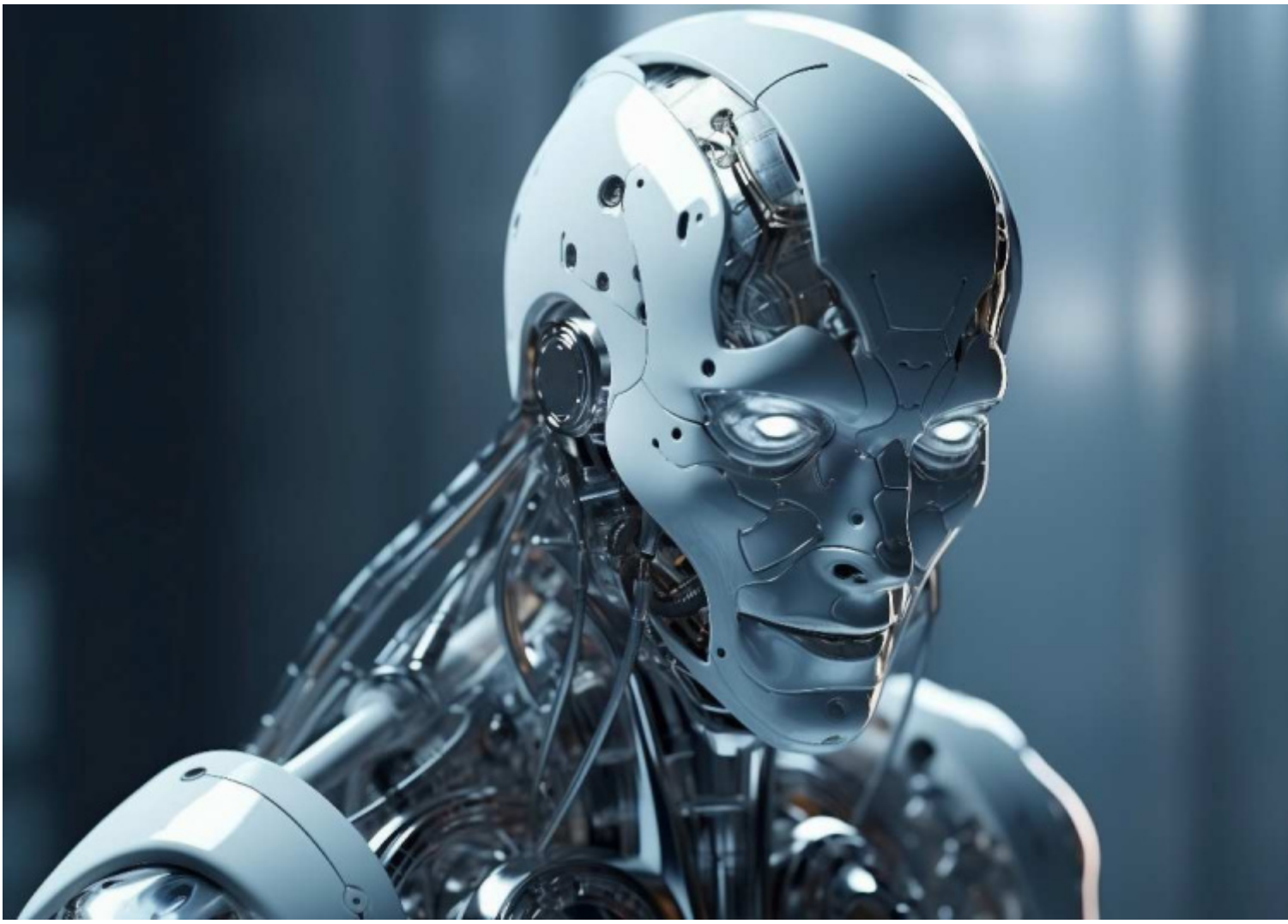
In questo dibattito il termine «coscienza» si riferisce soprattutto alla coscienza fenomenica, ovvero la capacità di sentire qualcosa; quella cosa che scompare (grazie al cielo) quando, prima di un intervento chirurgico, ci sottopongono a una anestesia totale. La coscienza fenomenica è distinta dalla coscienza morale, ben rappresentata dal grillo parlante di Pinocchio, che sarebbe la comprensione del significato delle nostre azioni. Tanto per capirci, un bambino di pochi anni non ha ancora una coscienza morale, ma sicuramente ha la coscienza fenomenica e infatti se si fa male lo sente benissimo! Una macchina, no. Ma perché no?

La coscienza è tutt’ora uno dei più grandi misteri irrisolti della scienza, tanto che in un sondaggio rivolto agli autori della rivista *Science*, è stata posta fra le principali domande senza una risposta insieme alle origini dell’universo e della vita. Oggi però, a seguito dello sviluppo dell’intelligenza artificiale, la coscienza non è

più solo un problema scientifico, ma è anche una sfida tecnologica. Oggi non è più velleitario chiedersi in concreto che differenza ci sia, in fondo, tra una macchina intelligente come GPT e una persona? Come ha affermato in una recente intervista Antonio Chella, professore di robotica a Palermo e da anni studioso di coscienza artificiale, «La coscienza è ciò che ci rende umani e, pertanto, per costruire un robot simile all'uomo, è inevitabile affrontare il problema della coscienza.»

È innegabile che fino a ieri la coscienza fosse associata, negli esseri umani, alle capacità logiche e soprattutto linguistiche. Nella pratica si attribuiva la coscienza alle creature dotate di linguaggio. Era molto comodo il fatto che gli esseri umani fossero l'unica specie in grado di parlare e scrivere. Nel caso degli animali si procedeva sulla base di una somiglianza anatomico-comportamentale con gli esseri umani: a gatti e scimmie si attribuiva qualche grado di coscienza alle meduse e alle aragoste no. Ma perché? Non si sa bene.

Le regole sono cambiate. GPT sa usare linguaggio e conoscenza e quindi il mistero della coscienza non è più all'interno di una rassicurante cornice antropocentrica. Quali sono le condizioni che consentono a un sistema fisico di passare il Rubicone esperienziale e diventare capace di provare qualcosa? Il problema della coscienza si presenta così diviso in due lati: da una parte il pericolo che macchine dotate di coscienza perseguano una autonomia che potrebbe non essere a nostro vantaggio e dall'altra il problema etico di gestire delle macchine nel momento in cui, in quanto dotate di sensazioni e motivazioni, non sarebbero più macchine.



Oggi nessuno studioso pensa che le intelligenze artificiali siano dotate di coscienza. Nessuno ritiene che quando digitiamo una richiesta a ChatGPT, questo sia in grado di fare esperienza delle nostre richieste. Tuttavia non sappiamo perché! Nessuno è in grado di fornire un criterio oggettivo per dire se un certo processo fisico (sia esso un'attività neurale dentro un organismo biologico o un processo elettronico dentro

un computer) sia cosciente. La coscienza è riuscita a sfuggire a ogni tentativo di spiegazione naturalistica. Fortunatamente abbiamo a disposizione una prova di esistenza: gli esseri umani sono coscienti e quindi la coscienza è possibile nel mondo fisico.

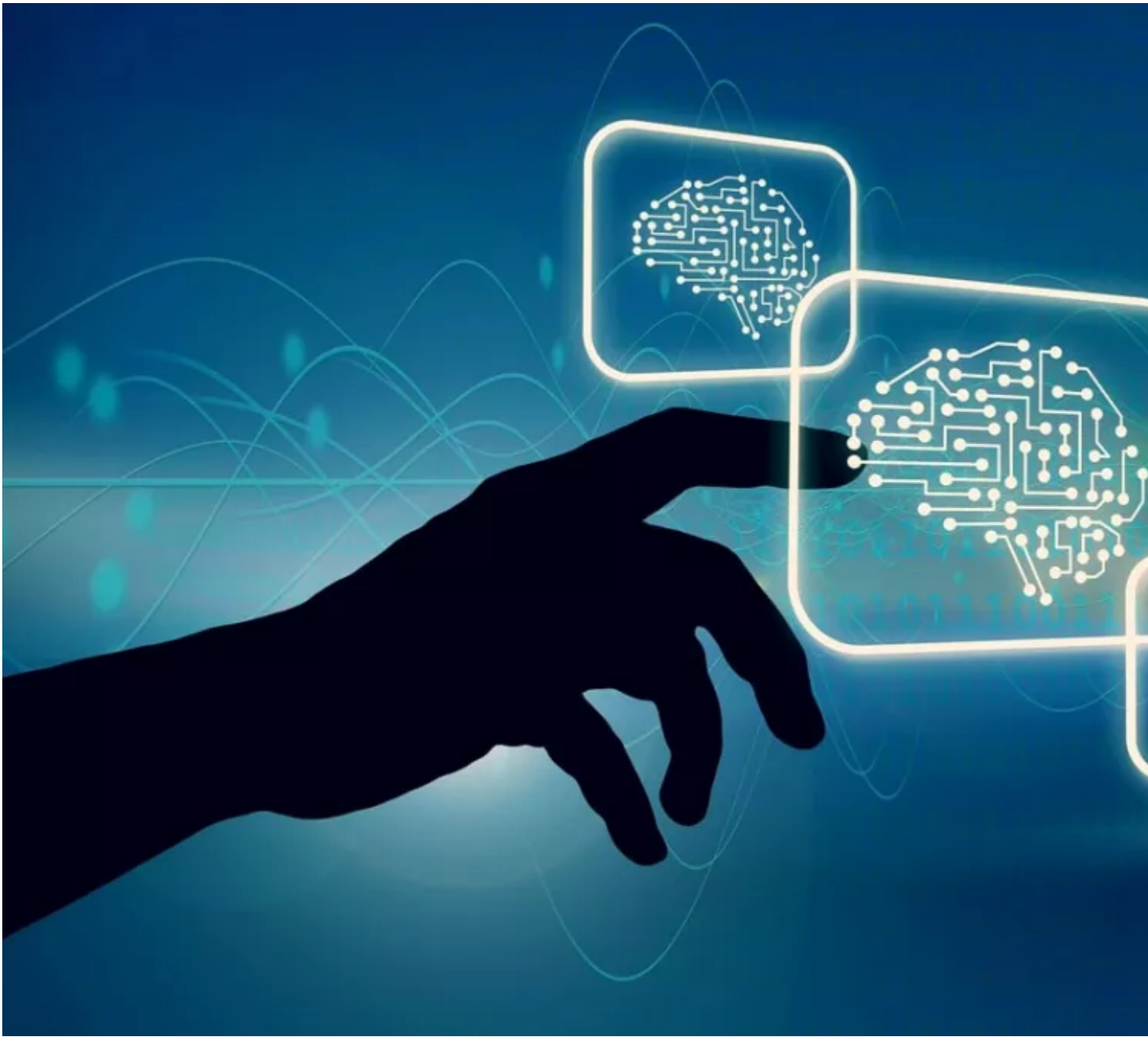
Nel caso dei modelli di linguaggio di grandi dimensioni, la tentazione di attribuire loro la coscienza è forte perché manipolano il linguaggio e la conoscenza a esso associata come un essere umano dotato di coscienza. Questi modelli usano il linguaggio con una competenza che, se fosse il prodotto di una persona, sarebbe sicuramente associata alla coscienza. È sufficiente per attribuire loro l'esperienza cosciente? Non pare, per quanto i motivi addotti non sembrino convincenti. Per esempio, spesso si sottolinea come queste macchine non contengano niente di simile alla mente, ma solo parametri statistici che rappresentano la probabilità che una parola sia usata in un certo contesto. L'obiezione ha un certo peso, ma diventa debole nel momento in cui si osserva che potrebbe essere ugualmente applicata agli esseri umani. Infatti, dentro il sistema nervoso non c'è alcunché di cosciente. Per quanto ne sappiamo, ci sono solo neuroni che, grazie alle loro connessioni, rappresentano la probabilità che una certa risposta segua a un determinato input. Insomma, quando formuliamo una frase, il nostro cervello, da un punto di vista fisico e matematico, non è poi così diverso da una intelligenza artificiale GPT. Il pensiero rimane qualcosa di misterioso ed evanescente e nessun neuroscienziato oggi può dire di averne visto uno, anche impiegando le tecniche di brain imaging più sofisticate.

Tuttavia, GPT ha dei limiti rispetto a una persona e cercherò di chiarirli facendo appello a un'analogia con la memoria umana. La memoria viene normalmente divisa in memoria semantica ed episodica. La memoria semantica è quella capacità che ha a che fare con la conoscenza linguistica (non a caso …). Per esempio, io mi ricordo che Cesare è stato assassinato il 44 a.C. alle idi di Marzo. A questo punto devo confessare che al momento non ho la più pallida idea di cosa siano le «idi». Il liceo è lontano. Ma se mi chiedete di formulare una frase sulla fine del divo Giulio, non ho esitazione nell'usare quell'espressione linguistica. Ecco, questo è un caso di conoscenza linguistica che non si associa ad alcuna esperienza diretta delle parole utilizzate. In molti casi, chi usa il linguaggio si trova in questa situazione. Si sa come usare le parole, ma questo avviene senza che tale pratica si accompagni alla coscienza in prima persona. La memoria episodica è invece sempre associata all'essere stati presenti a un certo evento, oggetto o situazione. Per esempio, se vi chiedo dove eravate quando vi siete sposati avete un ricordo personale. Muovendosi dal mondo della memoria a quello del linguaggio è come se le attuali intelligenze artificiali avessero solo una conoscenza semantica ma non episodica del mondo. È come se la coscienza artificiale fosse una specie di Emilio Salgari che, non solo non avesse mai visitato l'India, ma non fosse mai uscito da una stanza ermeticamente chiusa.

Nell'essere umano si procede dalla conoscenza diretta delle cose (o esperienza cosciente) verso la conoscenza semantica (o capacità di fornire descrizioni linguistiche). Il bambino *prima* vede e tocca oggetti che hanno colori, forme e sapori diverse e *dopo* impara il linguaggio che gli permette di comunicare. L'intelligenza artificiale si muove nella direzione opposta. *Prima* assimila i testi che le abbiamo messo a disposizione e *dopo* dovrebbe sviluppare la coscienza. Ma non si capisce perché dovrebbe e come potrebbe. Ha poco senso chiedersi se una intelligenza artificiale possa derivare dalla competenza linguistica il significato e quindi l'esperienza cosciente che, negli esseri umani, è stata la causa prima del linguaggio e della cultura. Personalmente, credo di no.

Non sempre le strade sono percorribili in entrambe le direzioni e ci sono molti motivi per pensare che il processo che va dal mondo al linguaggio sia a senso unico. Anche se nel secolo appena trascorso (il secolo breve secondo la fortunata definizione dello storico Eric Hobsbawm) ha avuto molto successo una visione trascendentale del linguaggio che lo ha posto come il succedaneo delle categorie kantiane. Al contrario credo che oggi si possa e si debba, anche grazie alla intelligenza artificiale, ripensare la realtà in termini *cosali*. Non è vero che il linguaggio sia la casa dell'essere, piuttosto il linguaggio è stato (ed è) una leva su cui le *cose* si appoggiano per produrre effetti nel mondo; nel nostro mondo che chiamiamo esperienza. L'intelligenza artificiale potrebbe essere l'occasione ideale per rimettere in discussione il rapporto tra mente e linguaggio, tra pensieri e proposizioni, tra esperienza e realtà. Grazie al fatto che la intelligenza artificiale sa usare il linguaggio in assenza di coscienza, potremmo concludere che la coscienza fa uso e si arricchisce del

linguaggio, ma non nasce e non coincide con esso.

Rimane l’imperativo del titolo di *Nature*: visto che l’intelligenza artificiale è così avanzata, non dovremmo essere in grado di distinguere fra sistemi coscienti e sistemi non coscienti? Non dovremmo essere in grado di distinguere chi è persona da chi non è? Come diceva il principe di Danimarca «Essere, o non essere, questo è il dilemma». Con buona pace del dualismo antropocentrico, essere è tutt’uno con l’essere coscienti. Il giorno che le macchine saranno coscienti, per la prima volta, oltre a noi, ci sarà qualcun altro che esisterà. La terra non sarà più solo nostra.

Se continuiamo a tenere vivo questo spazio è grazie a te. Anche un solo euro per noi significa molto.
Torna presto a leggerci e SOSTIENI DOPPIOZERO