Convivere con le macchine

Le sfide reali che l'IA solleva

Luciano Floridi

editoriale

"Il punto non è che le nostre macchine siano coscienti, intelligenti o capaci di conoscere qualcosa come noi. Non lo sono. Un sacco di macchine possono fare cose incredibili, come giocare a dama, scacchi e Go o al quiz show Jeopardy! meglio di noi. Eppure, sono tutte versioni di una 'macchina di Turing', un modello astratto che fissa i limiti di ciò che può essere realizzato da un computer tramite la sua logica matematica". Per questo motivo "dovremmo accendere la luce nella stanza buia e guardare attentamente dove stiamo andando. Non ci sono mostri ma molti ostacoli da evitare, rimuovere o negoziare". Occorre un'esplosione di intelligenza umana.

Negli anni Sessanta Irving John Good, un matematico britannico che ha lavorato come crittologo a Bletchley Park con Alan Turing, ha fatto la seguente osservazione: "Definiamo come ultraintelligente una macchina che può superare di gran lunga tutte le attività intellettuali di qualsiasi uomo per quanto intelligente. Poiché il design di macchine è una di queste attività intellettuali, una macchina ultraintelligente potrebbe disegnare macchine ancora migliori; assisteremmo allora indubbiamente a una 'esplosione d'intelligenza', mentre l'intelligenza umana sarebbe lasciata molto indietro. Per questo, la prima macchina ultraintelligente è l'ultima invenzione che l'uomo abbia bisogno di realizzare, a patto che la macchina sia sufficientemente docile da dirci come tenerla sotto controllo. È curioso che questo punto venga sollevato così di rado al di fuori della fantascienza. A volte, vale la pena di prendere sul serio la fantascienza".

Macchine ultraintelligenti

Una volta che le macchine ultraintelligenti diventassero realtà, potrebbero non essere affatto docili ma comportarsi come Terminator o meglio Skynet (nel film questa è la rete di difesa di IA che diventa autocosciente e dà inizio a un olocausto nucleare). Potrebbero schiavizzare l'umanità come specie inferiore, ignorarne i diritti e perseguire i propri fini, indipendentemente dagli effetti che ciò ha sulle vite umane. Se ciò sembra troppo incredibile per essere preso sul serio, il rapido avanzamento di mezzo secolo e gli incredibili sviluppi nelle nostre tecnologie digitali hanno portato alcune persone a credere che l'"esplosione d'intelligenza" di Good, talora presentata anche come "singolarità", possa costituire un serio rischio, e che, se non stiamo attenti, la fine della nostra specie potrebbe essere vicina. Nelle parole di Stephen Hawking: "Penso che lo sviluppo dell'Intelligenza Artificiale completa potrebbe dichiarare la fine della razza umana".

Bill Gates è parimenti preoccupato. Durante una sessione di domande e risposte "chiedimi qualsiasi cosa" su Reddit, ha scritto: "Sono tra quelli che si preoccupano per la superintelligenza. In un primo momento, le macchine faranno molti lavori per noi e non saranno superintelligenti. Ciò dovrebbe essere positivo se lo gestiamo bene. Pochi decenni dopo, tuttavia, l'intelligenza sarà abbastanza forte da destare preoccupazioni. Sono d'accordo con Elon Musk e altri su questo punto e non capisco perché alcune persone non se ne preoccupino".

E che cosa ha detto Elon Musk, esattamente? "Penso che dovremmo stare molto attenti all'Intelligenza Artificiale. Se dovessi scommettere su quale sia la nostra più grande minaccia esistenziale, direi che è probabilmente quella. Perciò, dobbiamo stare molto attenti. Un numero crescente di scienziati ritiene che ci dovrebbe essere una supervisione normativa, magari a livello nazionale e internazionale, che si assicuri che non facciamo qualcosa di davvero sciocco. Con l'Intelligenza Artificiale stiamo evocando il demone".

Negli ultimi anni Musk ha lanciato allarmi sempre più preoccupati. L'hanno seguito autori che hanno reso popolare la paura di una sorta di ultraintelligenza artificiale o superintelligenza. Molti non sono d'accordo, o semplicemente non prendono sul serio tali speculazioni. Alcuni li prendono in giro. Nel 2016, l'Information Technology and Innovation Foundation ha attribuito il suo annuale Premio Luddista (Luddite Award) a un'ampia coalizione di scienziati e luminari che nel 2015 ha suscitato paura e isteria lanciando allarmi sul fatto che l'Intelligenza Artificiale potrebbe segnare il destino dell'umanità. "È profondamente spiacevole che luminari come Elon Musk e Stephen Hawking abbiano contribuito alla delirante preoccupazione al riguardo di un'incombente apocalisse dell'Intelligenza artificiale", ha affermato il presidente dell'Itif Robert D. Atkinson.

La realtà è più banale e, in un certo senso, solleva preoccupazioni più realistiche. Le attuali e prevedibili tecnologie smart hanno l'intelligenza di un abaco, ossia zero.

Si può stare sereni?

[…] Il problema è sempre la stupidità umana o la natura malvagia. Pochi mesi dopo il Premio Luddista di cui sopra, il 23 marzo 2016, Microsoft ha introdotto Tay su Twitter: un chatbot basato sull'IA. Ha dovuto essere rimosso solo sedici ore dopo. Tay sarebbe dovuto diventare sempre più intelligente mentre interagiva con gli umani. Invece, è diventato rapidamente un malvagio sostenitore di Hitler, negazionista dell'Olocausto, promotore dell'incesto e assertore del fatto che "Bush ha realizzato l'11 settembre". Come mai? Perché funzionava come la carta assorbente da cucina, impregnandosi e assumendo la forma dei messaggi ingannevoli e sgradevoli che gli venivano inviati. Microsoft si è scusata.

Questo è lo stato dell'IA oggi e nel futuro realisticamente prevedibile. Ma non è un buon motivo per stare sereni. Al contrario, dopo tante speculazioni fuorvianti sui rischi inverosimili delle macchine ultraintelligenti, è giunto il momento di accendere la luce, smettere di preoccuparsi di scenari fantascientifici che distraggono e iniziare a concentrarsi sulle reali sfide dell'IA, per evitare di fare errori dolorosi e costosi nel design e nell'uso delle tecnologie smart.

I credenti nella "vera IA" e nell'"esplosione di intelligenza" di Good appartengono alla chiesa dei fautori della singolarità. In mancanza di un termine migliore, mi riferirò ai miscredenti come membri della chiesa degli atei dell'IA.

Gli adepti della singolarità credono in tre dogmi. Primo, la creazione di una qualche forma di ultraintelligenza artificiale è probabile o almeno non impossibile nel (per alcuni di loro prevedibile) futuro. Questa svolta è nota come "singolarità tecnologica", da cui il nome. Sia la natura di tale superintelligenza sia l'esatto lasso di tempo del suo arrivo non sono specificati, sebbene gli adepti della singolarità tendano a preferire futuri che sono convenientemente abbastanza vicini da preoccuparsene ma sufficientemente lontani da non essere più là per verificarne la correttezza o no. In secondo luogo, l'umanità corre il grosso rischio di essere dominata da tale ultraintelligenza. In terzo luogo, la generazione attuale ha la responsabilità primaria di assicurare che la singolarità non abbia luogo o, se accade, che sia benigna e vada a vantaggio dell'umanità. Ciò ha tutte le caratteristiche di una visione manichea del mondo: il Bene che combatte il Male, toni apocalittici, l'urgenza del "dobbiamo fare qualcosa ora o sarà troppo tardi", una prospettiva escatologica della salvezza umana, e un appello alle paure e all'ignoranza.

[…] Profondamente irritati da coloro che adorano gli dei digitali sbagliati e dalle loro profezie irrea-

lizzate sulla singolarità, i miscredenti (gli atei dell'IA) intraprendono la missione di dimostrare una volta per tutte che qualsiasi tipo di fede nella vera IA è sbagliata, del tutto sbagliata. L'IA è solo computer, i computer sono solo "macchine di Turing", le macchine di Turing sono solo motori sintattici e i motori sintattici non possono pensare, non possono sapere, non possono essere coscienti. Fine della storia.
Questo è il motivo per cui ci sono così tante cose che i computer non possono ancora fare, anche se ciò che esattamente non possono fare è un bersaglio che può essere opportunamente spostato. È anche il motivo per cui non sono in grado di processare la semantica (di qualsiasi lingua, incluso il cinese), indipendentemente da ciò che la traduzione di Google riesce a ottenere. Ciò dimostra che non c'è assolutamente nulla di cui discutere, e tantomeno di cui preoccuparsi. Non esiste una vera IA, e per questo non ci sono a fortiori problemi causati da essa. Possiamo rilassarci e goderci tutti questi meravigliosi dispositivi elettrici.

Il rischio di un dibattito inutile

La fede degli atei è malriposta quanto quella degli adepti della singolarità. Entrambe le chiese hanno molti seguaci in California, dove i film di fantascienza di Hollywood, le meravigliose università di ricerca come Berkeley e alcune delle aziende digitali più importanti del mondo prosperano fianco a fianco. Ciò può non essere un caso. Quando ci sono molti soldi in gioco, le persone si confondono facilmente.
[…] Sia gli adepti della singolarità sia gli atei dell'IA si sbagliano. Come Turing ha affermato chiaramente nel 1950, la domanda "una macchina è in grado di pensare?" è "troppo insignificante per meritare una discussione". Eppure, entrambe le chiese proseguono questo inutile dibattito, soffocando spesso ogni voce dissenziente della ragione. La vera IA non è logicamente impossibile, ma è assolutamente non plausibile. Le persone confondono "la singolarità non accadrà mai" con "la singolarità è impossibile". Impossibile è un concetto logico e la vera IA è logicamente possibile. Ma è possibile come un calcolo che, per esempio, richiederebbe più tempo della vita dell'universo per essere completato: non accadrà.

Quello che siamo, quello che le macchine non sono

Ciò che conta davvero è che la presenza crescente di tecnologie sempre più smart nelle nostre vite sta avendo un enorme impatto sul modo in cui concepiamo noi stessi, il mondo e le interazioni che intratteniamo tra noi e con il mondo. Il punto non è che le nostre macchine siano coscienti, intelligenti o capaci di conoscere qualcosa come noi. Non lo sono. Un sacco di macchine possono fare cose incredibili, come giocare a dama, scacchi e Go o al quiz show Jeopardy! meglio di noi. Eppure, sono tutte versioni di una "macchina di Turing", un modello astratto che fissa i limiti di ciò che può essere realizzato da un computer tramite la sua logica matematica. Anche i computer quantistici sono vincolati dagli stessi limiti, i limiti di ciò che può essere calcolato (le cosiddette funzioni computabili). Nessuno sembra capace di spiegare come un ente cosciente, intelligente ed empatico possa emergere da una macchina di Turing.

Il punto è che le nostre tecnologie smart, anche grazie all'enorme quantità di dati disponibili e a programmi molto sofisticati, sono sempre più capaci di svolgere un numero crescente di compiti meglio di noi, compresa la previsione dei nostri stessi comportamenti, senza dover essere affatto intelligenti. Per questo, non siamo gli unici agenti in grado di svolgere compiti con successo, tutt'altro. È quello che ho definito "Quarta rivoluzione" nella comprensione di noi stessi. Non siamo al centro dell'universo (Copernico), del regno biologico (Darwin) o del regno della razionalità (Freud). Dopo Turing, non siamo più al centro dell'infosfera né del mondo dell'elaborazione delle informazioni e dell'agire smart. Non ho mai sostenuto che le tecnologie digitali pensino meglio di noi, ma che possano fare sempre più cose meglio di come le facciamo noi senza pensare, limitandosi a elaborare quantità crescenti di dati in modo sempre più efficiente ed efficace.

editoriale

Artefatti ordinari

Le tecnologie digitali [...] sono artefatti ordinari che ci sopravanzano in un numero crescente di compiti, nonostante non siano più intelligenti di un tostapane. Le loro capacità sono umilianti e ci fanno riconsiderare la nostra eccezionalità umana e il nostro ruolo speciale nell'universo, che rimane unico. Pensavamo di essere intelligenti perché sapevamo giocare a scacchi. Ora un telefono gioca meglio di un maestro di scacchi. Pensavamo di essere liberi perché potevamo comprare quello che volevamo. Ora i nostri modelli di spesa sono previsti, a volte addirittura anticipati, da dispositivi stupidi come una zucchina. Che cosa significa tutto questo per la comprensione che abbiamo di noi stessi? Questa è una domanda che vale la pena di indagare dal punto di vista filosofico.

Ignorare le visioni apocalittiche

Il successo delle nostre tecnologie dipende in gran parte dal fatto che, mentre speculavamo sulla possibilità dell'ultraintelligenza, abbiamo sempre più avvolto il mondo per mezzo di così tanti dispositivi, sensori, applicazioni e dati da diventare un ambiente adattato alle Information and communication technologies (Ict), dove le tecnologie possono sostituirci senza disporre di alcuna comprensione, stato mentale, intenzione, interpretazione, stato emotivo, abilità semantiche, coscienza, autocoscienza o intelligenza flessibile. La memoria (come quella presente in algoritmi e immensi set di dati) supera l'intelligenza quando si tratta di far atterrare un aereo, individuare il percorso più veloce da casa all'ufficio o scoprire il prezzo migliore per il nostro prossimo frigorifero. Le tecnologie digitali possono fare sempre più cose meglio di noi, elaborando quantità crescenti di dati e migliorando le loro prestazioni, analizzando il proprio output come input per le operazioni successive.

Qualsiasi visione apocalittica dell'IA può essere ignorata. Il vero rischio non sta nella comparsa di qualche forma di ultraintelligenza, ma nel fatto che possiamo utilizzare male le nostre tecnologie digitali, a danno di una grande percentuale dell'umanità e dell'intero pianeta. Noi siamo e rimarremo, in qualsiasi prevedibile futuro, il problema, non la nostra tecnologia. Questo è il motivo per cui dovremmo accendere la luce nella stanza buia e guardare attentamente dove stiamo andando. Non ci sono mostri ma molti ostacoli da evitare, rimuovere o negoziare.

Dovremmo preoccuparci della vera stupidità umana, non dell'Intelligenza Artificiale immaginaria, e concentrarci sulle sfide reali che l'IA solleva.

Tratto da "Etica dell'Intelligenza artificiale. Sviluppi, opportunità, sfide", Raffaello Cortina Editore, Milano 2022, cap. X. Per gentile concessione dell'autore e di Raffaello Cortina Editore ©.

Luciano Floridi è un filosofo italiano naturalizzato britannico, professore ordinario di Filosofia ed etica dell'informazione presso l'Università di Oxford, dove è direttore del Digital Ethics Lab, nonché professore di Sociologia della comunicazione presso l'Università di Bologna.