

I RISCHI DELLA SUPERINTELLIGENZA

di GIACOMO CORVI

UN RECENTE STUDIO LANCIA L'ALLARME SULLE CONSEGUENZE DI UNA TECNOLOGIA FUORI CONTROLLO: NIENTE E NESSUNO SAREBBE IN GRADO DI FERMARE UN DISPOSITIVO CON CAPACITÀ COGNITIVE DI GRAN LUNGA SUPERIORI ALLE POSSIBILITÀ UMANE

Che cosa succederebbe se una superintelligenza fosse incaricata di realizzare il maggior numero possibile di graffette? La risposta, per quanto possa sembrare stupida la domanda, non è scontata. E affligge da anni il lavoro di scienziati, filosofi ed esperti di nuove tecnologie. Il quesito è stato formulato nel 2003 dal filosofo svedese **Nick Bostrom**, oggi direttore del *Future of Humanity Institute* presso l'**Università di Oxford**. E non ha ancora trovato una risposta definitiva. O meglio, una risposta forse c'è, ma non è sicuramente quella che ci aspetteremmo e vorremmo: incaricata di realizzare il maggior numero possibile di graffette, una *superintelligenza* potrebbe arrivare a sterminare l'umanità.

Il concetto di superintelligenza è diventato di moda negli ultimi anni, sulla scia dei sempre più sorprendenti traguardi raggiunti da algoritmi di intelligenza artificiale. Bostrom, nel suo best-seller *Superintelligenza. Tendenze, pericoli, strategie*, la definisce come "qualunque intelletto che superi di molto le prestazioni cognitive degli esseri umani in quasi tutti i domini di interesse". Si tratterebbe dunque di una capacità cognitiva di gran lunga superiore alle possibilità umane, in grado di trovare connessioni e nessi di causalità

dove noi vediamo solo caos. Per rendere l'idea, Bostrom paragona a più riprese il divario che esisterebbe fra superintelligenza e intelligenza umana a quello che sussiste oggi fra un essere umano e un verme. Solo che questa volta il verme siamo noi.

TUTTO PER DELLE GRAFFETTE

Frank Lantz, direttore del *Game Center* della **New York University**, ha creato nel 2017 una sorta di videogame basato sulla teoria di Bostrom. Il gioco, battezzato semplicemente *Paperclips* e tuttora disponibile online, ti mette nei panni di un'intelligenza artificiale intenta a realizzare graffette. Inizialmente bisogna soltanto stabilire un prezzo di vendita e cliccare ogni volta che si vuole produrre una singola graffetta. Poi diventa tutto più complesso: bisogna comprare macchinari sempre più efficienti, rifornirsi di materie prime, definire strategie di marketing, trovare nuove fonti di finanziamento e rilevare le attività dei concorrenti. La piega finale è del tutto inaspettata: viene introdotta la computazione quantistica, bisogna esplorare l'universo per trovare nuove materie prime e persino lottare contro gli alieni per difendere le proprie graffette. Nel mezzo c'è anche posto per quelli che vengono definiti



hypnodrones, una diavoleria in grado di ipnotizzare la popolazione per spingere al rialzo la domanda. Tutto per delle graffette.

L'escalation del gioco fa ben comprendere quanto veloce possa correre una superintelligenza. Per gli esperti si tratta addirittura di uno scenario al ribasso, visto che il gioco è stato sviluppato da un essere umano e una superintelligenza potrebbe invece prendere decisioni per noi inimmaginabili. Ecco perché è fondamentale definire fin da ora strumenti in grado di evitare conseguenze di questo tipo. Sempre ammesso che una superintelligenza possa essere controllata.

TECNOLOGIA FUORI CONTROLLO

Uno studio recente, pubblicato sul *Journal of Artificial Intelligence Research*, non sembra lasciare spazio a dubbi: una superintelligenza non potrà mai essere controllata. Il team di ricerca ha esplorato l'ipotesi di un algoritmo di contenimento in grado di interrompere l'attività della superintelligenza se sussistesse il rischio che possa danneggiare un essere umano. Ebbene, stando ai risultati della ricerca, un dispositivo di sicurezza di questo tipo è semplicemente impossibile. "Nel nostro attuale paradigma di calcolo, tale algoritmo non può essere costruito", ha commentato **Iyad Rahwan**, direttore del *Center for Humans & Machines* del **Max-Planck Institute for Human Development**, che ha collaborato alla ricerca. "Sulla base di questi calcoli – ha proseguito – il problema del contenimento

è incomputabile, ovvero nessun singolo algoritmo può trovare una soluzione per determinare se un'intelligenza artificiale potrebbe produrre danni al mondo". Inoltre, quasi a voler incrementare ulteriormente l'ansia data da una possibile tecnologia fuori controllo, Rahwan ha sottolineato che, nonostante tutti gli sforzi, "potremmo anche non accorgerci del momento in cui le macchine superintelligenti emergeranno, perché stabilire il grado di intelligenza di un sistema rientrerebbe nello stesso ambito di analisi".

UN RISCHIO IGNORATO

A dire il vero, una soluzione ci sarebbe: una superintelligenza potrebbe infatti essere deliberatamente limitata nelle sue funzioni, isolata dal resto del mondo e resa incapace di apprendere oltre le capacità di controllo di noi esseri umani. Così facendo, però, non avremmo una vera superintelligenza. E tutti i problemi per cui l'umanità potrebbe aver interesse a realizzare una superintelligenza, come curare malattie o ridurre la disuguaglianza, resterebbero pertanto irrisolti.

In questo contesto, l'intera umanità si trova di fronte a un bivio: correre il rischio di sviluppare una tecnologia incontrollabile o convivere per sempre con problemi che finora non siamo riusciti a risolvere. La questione non è assolutamente banale. E negli anni ha attirato l'attenzione di numerosi esperti e addetti ai lavori: da **Bill Gates** a **Elon Musk**, fino ad arrivare allo scienziato britannico **Stephen Hawking**, il quale negli ultimi anni della sua vita si era molto soffermato sui possibili rischi di un'intelligenza artificiale fuori controllo. "Un'intelligenza artificiale superintelligente sarà estremamente brava a realizzare i propri obiettivi, e se questi obiettivi non si allineeranno con i nostri, saremo nei guai", aveva detto poco prima della sua scomparsa. Un monito che, almeno per il momento, sembra caduto nel vuoto. La pandemia di coronavirus ha dominato le tradizionali classifiche di inizio anno sui rischi a livello globale: nel 2021, com'è forse inevitabile dopo un anno di emergenza sanitaria, spaventano soprattutto pandemie, interruzione di attività e crimini informatici. Della superintelligenza, o comunque di eventuali conseguenze infauste delle nuove tecnologie, non c'è alcuna traccia ai vertici delle graduatorie. ❶