



L'Intelligenza artificiale è già in grado produrre studi e articoli di ricerca (con contenuti potenzialmente falsi) indistinguibili da quelli veri. E nemmeno gli esperti del settore sono in grado di distinguere tra algoritmo e mente umana.

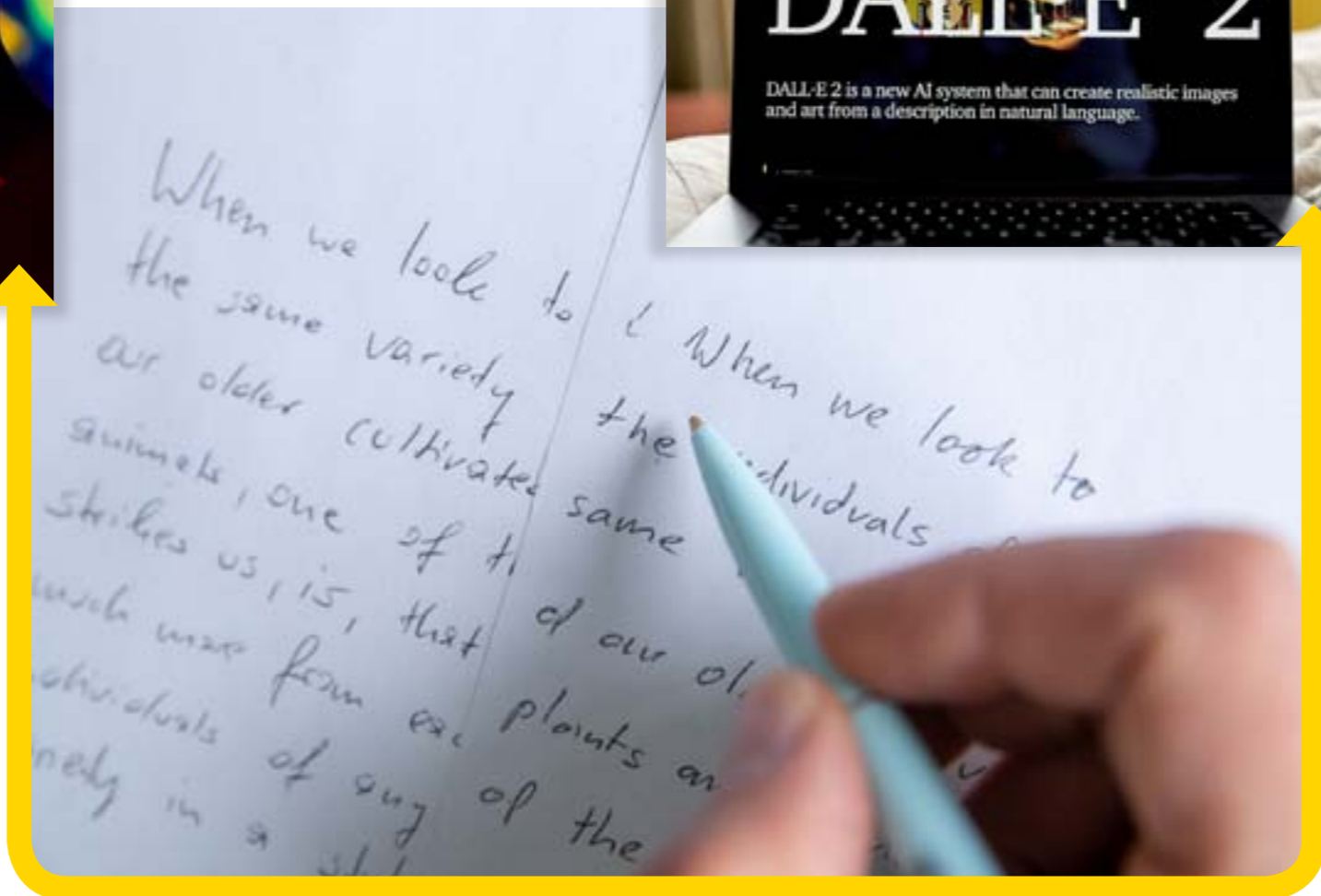
di Luca Sciortino

C'è articolo e articolo. Uno come questo si propone di informare rivolgendosi a un largo pubblico. Altri, pubblicati da riviste specializzate, sono indirizzati a esperti in un dato ambito della ricerca e comunicano avanzamenti della conoscenza giustificandone la plausibilità attraverso metodi e procedimenti oggettivi.

È difficile sottovalutare l'importanza di questi ultimi. Costituiscono il perno di decisioni politiche, sociali, sanitarie, educative con enormi ripercussioni sulla vita di ognuno di noi. Proprio per questo, preoccupa il fatto che l'intelligenza artificiale, con i suoi nuovi programmi di analisi, revisione e riprogettazione di testi, sia ormai in grado di produrre finti studi capaci di ingannare gli scienziati stessi: all'apparenza, questi pseudo-articoli seguono metodi riconosciuti nel campo,

hanno una logica interna e fanno un uso corretto della terminologia e della letteratura sull'argomento; ma di fatto non costituiscono un reale progresso della conoscenza scientifica e possono perfino essere scorretti nelle conclusioni. L'ultimo grido di allarme arriva da *Nature*, la rivista regina di questo tipo di articoli: «Il programma ChatGBT crea testi realistici e razionalmente accettabili in risposta a precisi suggerimenti degli studiosi. Da quando la compagnia di software californiana Open AI l'ha messo in commercio, sono emersi tra i ricercatori innumerevoli problemi etici riguardanti il suo uso, perché i testi prodotti sono quasi indistinguibili da quelli elaborati da una mente umana».

Qualche mese fa alcuni scienziati avevano generato per prova, tramite ChatGPT, un articolo di ricerca vero e proprio e anche un editoriale di carattere scientifico, all'ap-



Lo scienziato è un robot (ma nessuno lo sa)

parenza credibili. Adesso, un gruppo di studiosi della Northwestern University di Chicago è andato oltre fornendo una vera e propria valutazione di come è possibile produrre cosiddetti «abstract» (sommari di articoli di ricerca) capaci di ingannare gli esperti.

Al programma veniva richiesto di scrivere 50 di questi brevi testi basandosi su una selezione di altri abstract originati dalla mente umana per riviste autorevoli quali *Jama*, *New England of Medicine*, *The Bmj*, *The Lancet* e *Nature Medicine*.

Algoritmi come ChatGPT apprendono grazie all'esperienza, proprio come il bambino impara a riconoscere un gatto dopo che i suoi genitori gliel'hanno mostrato più volte. In questo caso viene fornito al computer come dato iniziale un «insieme di addestramento», cioè una serie di abstract, dai quali impara la struttura, la terminologia, l'ordine delle informazioni e la sua contestualizzazione. Più l'insieme è ampio più l'algoritmo del programma risulterà efficiente.

Il gruppo dell'Università

A lanciare l'allarme su studi e articoli di ricerca prodotti da algoritmi avanzati come ChatGPT o DALL-E 2 (a sinistra) è la rivista *Nature*, tra le più autorevoli in campo scientifico.

di Chicago ha prima sottoposto gli abstract costruiti da ChatGPT a un altro programma usato dalle riviste scientifiche per identificare il plagio. Risultato: i testi apparivano originali, caratteristica cruciale nella ricerca. Infine, gli abstract venivano esaminati da differenti revisori, cioè studiosi incaricati di valutare anonimamente la correttezza degli articoli. Questi ultimi hanno identificato come falsi solo il 68 per cento dei 50 pseudo-abstract e, di fronte a un analogo campione di articoli genuini scritti da umani, affermavano - erroneamente - che il 14 per cento era generato dall'A.I.

Quindi, ora sappiamo che 32 su 100 articoli prodotti da un'intelligenza artificiale, e potenzialmente contenenti informazioni false sulla medicina e la salute, passerebbero indenni il vaglio della peer review (la revisione nell'anonimato di un articolo).

Ciò non toglie che al procedere delle ricerche non ci si possa accorgere se un articolo è scritto da un computer, ma certo siamo di fronte a un forte «fattore di disturbo» all'interno del processo scientifico di un tipo finora sconosciuto nella storia umana.

Nella ricerca sarà più difficile distinguere il vero dal falso. Hafteinn Einarsson, ingegnere informatico dell'Università di Reykjavic dice:

«Esiste tutta un'area dell'intelligenza artificiale che aiuta i computer a produrre testi perfettamente leggibili. Diversi programmi assistono nella ricerca, io ne uso uno che si chiama GPT-3».

Si tratta di un programma che genera testi a partire da suggerimenti o domande con un costo di 0,0004 dollari ogni 750 parole. «Mi consente di scrivere meglio una parte di testo, fare un titolo, individuare parole chiave, riassumere un articolo e molto altro. Insomma, mi fa risparmiare tempo e fatica» conclude Einarsson.

Tra gli algoritmi che aiutano gli scienziati si potrebbe citare il cosiddetto AI2, che «ingurgita» un intero articolo di ricerca e poi risponde a differenti domande sul suo contenuto; o l'ACCoRD, capace di generare definizioni e analogie di concetti scientifici.

Nemmeno le discipline umanistiche sono immuni. Le «digital humanities» sono un'area in cui l'A.O. viene usata per analizzare testi. Tra le cose mirabili che si possono fare c'è quella di rintracciare l'uso di un termine nell'opera omnia di un autore facendone comparazioni allo scopo di estrarne un'interpretazione. Che un giorno comprenderemo un libro e non saremo in grado di distinguere se l'ha scritto un umano o una macchina è solo questione di tempo. ■