

Ma gli androidi scrivono saggi elettrici?

Intelligenza artificiale e pensiero umano. Filosofia per un tempo nuovo
di John R. Searle.

di [Luca Bandirali](#) – 27 Febbraio 2023



L'irruzione sul mercato di nuovi software *chatbot* basati sul modello GPT-3, in grado di dialogare con utenti e anche di produrre testi originali, sta generando l'ennesimo dibattito sull'intelligenza artificiale, dibattito puntualmente divisivo in cui le fazioni si organizzano in apocalittici e integrati. Le macchine scriveranno tesi di laurea al posto degli studenti? Pubblicheranno saggi e monografie? Diffonderanno notizie false? E, alla fine, superate con profitto tutte le abilitazioni, assoggetteranno l'umanità come in *Matrix*? Oppure, al contrario, i software di questa generazione ci aiuteranno genericamente a vivere meglio? **Tra le voci del dibattito, va citata una frase di Bill Gates, fondatore ed ex-CEO di Microsoft, secondo il quale la differenza tra le applicazioni precedenti dell'intelligenza artificiale e quelle basate sul modello GPT-3 sta nel fatto che ora le macchine sono in grado di «comprendere il contenuto»** (ha detto letteralmente «understand the content»; la fonte è un'intervista pubblicata sul quotidiano tedesco "Handelsblatt").

Questa è una frase che ci proietta subito nella Stanza Cinese, uno degli esperimenti mentali più noti della filosofia contemporanea. Congegnato da John R. Searle, l'esperimento della Stanza Cinese consiste nell'immaginare una persona che non conosce la lingua cinese e che si trova in una stanza chiusa, senza possibilità di comunicare con l'esterno, ma ha a disposizione un elenco di ideogrammi e un testo di istruzioni in inglese che spiega con quale sequenza di simboli si risponde quando si riceve una determinata combinazione. Da sotto la porta, riceve delle domande scritte in cinese e, pur non conoscendo la lingua, le sue risposte realizzate sulla base delle istruzioni hanno senso; pertanto, l'interlocutore esterno è convinto che l'ospite della stanza sappia il cinese. Searle conclude, nel suo celebre saggio pubblicato nel 1980: «I have inputs and outputs that are indistinguishable from those of the native Chinese speaker, [...], but I still understand nothing».

L'esperimento mentale di Searle dimostra che il computer può parlare in cinese, certamente, ma senza capirlo: la persona nella stanza, cioè il computer, non capisce nulla. **Ora, aveva ragione Searle nel 1980 o ha ragione Bill Gates nel 2023? O hanno ragione entrambi, nel senso che le vecchie intelligenze artificiali non erano così intelligenti, mentre il GPT-3 consente finalmente alle macchine di «comprendere il contenuto»?**

Una risposta degna di interesse è offerta dall'ultimo libro di Searle, intitolato *Intelligenza artificiale e pensiero umano. Filosofia per un tempo nuovo*, curato da Angela Condello per Castelvecchi. Il volume presenta, per la prima volta in lingua italiana, sei saggi pubblicati su rivista tra il 1990 e il 2006, testi che fanno il punto sulle questioni più rilevanti della ricerca del filosofo americano: ontologia sociale, atti linguistici, intenzionalità e intelligenza artificiale. In appendice, un dialogo molto recente del filosofo con la curatrice Angela Condello ribadisce le posizioni di Searle sull'attuale impossibilità di duplicazione della coscienza nonostante i progressi della tecnologia computazionale.

Secondo Searle, tutto il dibattito sulla relazione tra umano e macchinico è minato alla base da una serie di errori filosofici, fra i quali ci sono il dualismo mente/corpo e la confusione tra simulazione e duplicazione. Il dualismo mente/corpo, che Searle ritiene obsoleto alla luce delle nostre conoscenze attuali sul cervello umano e animale, considera i fenomeni mentali differenti e separati dai fenomeni biologici, mentre invece sono la stessa cosa: i fenomeni mentali sono biologici come la fotosintesi e la digestione. Come questi ultimi possono essere duplicati, così, in linea di principio, anche i fenomeni mentali potrebbero essere riprodotti; ma c'è una differenza, questa sì sostanziale, tra duplicare un fenomeno e simularlo. **Un cuore artificiale non simula il funzionamento del cuore umano, ma pompa effettivamente sangue; la cosiddetta**

intelligenza artificiale, al contrario, si limita a simulare la coscienza e non è attualmente in grado di produrla.

L'aspetto simulatorio dell'intelligenza artificiale è perfettamente chiarito dall'esperimento della Stanza Cinese; come ribadisce Searle nell'intervista che chiude il volume (datata ottobre 2022), la sintassi (ossia la manipolazione di simboli in base a regole) «non è sufficiente a costituire il tipo di contenuto semantico di una mente cosciente reale». **La macchina svolge sempre più operazioni, ma «i progressi della tecnologia computazionale sono stati finora quantitativi, non qualitativi»: il computer fa più cose di prima e le fa più velocemente, ma non capisce nulla di ciò che fa.**

La risposta di Searle, che non è certo un apocalittico, non è definitiva: per duplicare un fenomeno abbiamo bisogno di sapere perfettamente come funziona, ma nel caso del cervello umano e animale sono i poteri causali a determinare la coscienza, e tali poteri ci sono ancora «largamente ignoti». Insomma, un *chatbot* potrà anche aver scritto questo articolo, ma sicuramente non lo ha pensato.

John R. Searle, *Intelligenza artificiale e pensiero umano. Filosofia per un tempo nuovo*, Castelvecchi, Roma 2023.