

Mente e computer: la filosofia tra svolta cognitiva e intelligenza artificiale

“ Qui assumo che il cervello sia un sistema fisico e che quindi obbedisca a leggi che possono essere simulate computazionalmente. Di sicuro, le prove attuali favoriscono entrambe le assunzioni. Se sono vere, allora dovrebbe essere possibile simulare al computer un cervello.

(David Chalmers, *Più realtà*)

1. I processi cognitivi e l'intelligenza artificiale

■ La svolta cognitiva della filosofia analitica

Il rovesciamento del rapporto mente-linguaggio Nella concezione che la filosofia analitica aveva ereditato dai suoi fondatori – Frege, Russell, Wittgenstein e Carnap – il pensiero e il linguaggio erano essenzialmente connessi, e il loro legame era tale per cui la **comprensione della mente** richiedeva l'**analisi del linguaggio**. Negli anni Ottanta del Novecento, all'interno della stessa filosofia analitica, questo principio generale viene messo in discussione. Decisivo in tal senso è il contributo di **Gareth Evans** (1946-1980), il quale in *The Varieties of Reference* (letteralmente “Le varietà del riferimento”, 1982), propone di concepire la relazione fra mente e linguaggio partendo dalla prima anziché dal secondo. Per Evans, infatti, a **fondamento del linguaggio** vi è la **capacità della mente** di fare **riferimento agli oggetti del mondo**, e questa capacità è essenzialmente pre-concettuale e pre-linguistica, ossia basata non su astratte strutture logiche, bensì sul fatto che abbiamo un corpo che ci permette di percepire e agire. Scrive Evans:

“ I concetti ai quali ricorriamo quando pensiamo sono appresi da un organismo che già dispone di connessioni fra gli input sensoriali e il comportamento.

(*The Varieties of Reference*, cap. 6)

Il ritorno allo studio dei processi psichici Mentre la tradizione analitica, a partire da Frege, aveva studiato la mente e il linguaggio in una prospettiva anti-psicologica, cioè facendo **astrazione dai processi psichici** in cui il pensiero si realizza, Evans ritiene che per comprendere la mente e il linguaggio sia indispensabile considerare proprio che cosa accade nella mente del soggetto, inteso come organismo capace di pensiero e di





linguaggio, ma soprattutto di percezione e azione. Si parla a questo proposito di una “**svolta cognitiva**” della filosofia analitica, che porta in primo piano le **operazioni psichiche** necessarie per la rappresentazione sia mentale sia linguistica del mondo.

■ Le scienze cognitive e il funzionalismo

Nel quadro della svolta cognitiva, vari filosofi analitici prendono parte al programma di ricerca delle **scienze cognitive**. L'espressione indica un insieme di **discipline** – dalla psicologia all'antropologia, dalla linguistica all'informatica, alle **neuroscienze** e alla stessa filosofia – che contribuiscono congiuntamente allo studio dell'attività mentale.

La mente come software Facendo leva sull'analogia fra il soggetto umano e il computer, le scienze cognitive studiano l'attività mentale come un processo di **acquisizione** e di **elaborazione di informazioni** reperibili nell'ambiente. La mente viene così concepita come un “software” il cui “hardware” è il corpo.

Il modello funzionalista Nella prospettiva delle scienze cognitive, il soggetto umano può essere paragonato a un computer che riceve *input* dall'ambiente e li elabora per produrre *output* che vanno a modificare l'ambiente stesso. Gli *input* sono innanzitutto forniti dalle **percezioni**, mentre gli *output* consistono nelle **azioni**.

Altri stati mentali – come le emozioni, le immaginazioni, i ricordi, le credenze, i desideri e le intenzioni – contribuiscono invece al funzionamento del “**sistema di elaborazione**”, il quale partendo dagli *input* percettivi produce gli *output* comportamentali.

scienze cognitive sono le scienze che studiano i processi cognitivi, ossia cercano di comprendere e descrivere gli stati e le operazioni mediante i quali la mente acquisisce, memorizza, rielabora e utilizza informazioni in vista dell'azione.

neuroscienze sono le diverse scienze che concorrono allo studio degli aspetti funzionali del sistema nervoso: dalla fisiologia alla biologia, dalla chimica alla psicologia, alla pedagogia e alla linguistica.

**LESSICO
FILOSOFICO**

In una prima fase, la percezione origina ricordi e credenze, a cui possono aggiungersi immaginazioni, emozioni e – in conseguenza di bisogni e istinti corporei – desideri; in una seconda fase, la combinazione di credenze e desideri produce le intenzioni, che infine conducono all'*output* dell'azione. Si parla anche, a questo proposito, di **funzionalismo**, nel senso che gli stati mentali sono definiti in base alla *funzione* che svolgono nel processo complessivo di elaborazione dell'informazione.

■ La teoria modulare di Fodor

Sempre in analogia con il computer, che è composto da vari dispositivi con funzioni specifiche, il filosofo americano **Jerry Fodor** (1935-2017) ha introdotto la nozione di «**modularità della mente**». La sua idea è che la vita mentale si basi su differenti «moduli» deputati a svolgere differenti funzioni. Ad esempio, Fodor ritiene che la percezione sia un modulo che funziona autonomamente rispetto ad altre forme più astratte di conoscenza, il che contribuisce a spiegare perché continuiamo a cadere vittima di certe **illusioni percettive**, pur sapendo che si tratta appunto di percezioni illusorie. Emblematica a questo proposito è l'illusione ottica descritta dallo psicologo tedesco Franz **Müller-Lyer** (1857-1916): se alle estremità di due segmenti di uguale lunghezza vengono aggiunti angoli differenti, un segmento ci appare più lungo dell'altro.



Nell'illusione di Müller-Lyer, continuiamo a *vedere* i due segmenti come aventi lunghezze diverse, pur *sapendo* che sono della stessa lunghezza. Nella prospettiva fodoriana, questo accade perché la percezione è un modulo che funziona per conto proprio.

■ Le origini dell'intelligenza artificiale

L'analogia fra mente e computer rende possibile un proficuo dialogo fra le scienze cognitive e un altro programma di ricerca avviato negli anni Cinquanta: l'**intelligenza artificiale**. Mentre le scienze cognitive cercano di comprendere la mente prendendo come modello esplicativo il computer, l'intelligenza artificiale procede in direzione opposta: si avvale degli studi sulla mente per programmare i computer in modo che possano svolgere un'attività cognitiva paragonabile a quella degli esseri viventi.

La macchina di Turing Pioniere dell'intelligenza artificiale è considerato il matematico inglese **Alan Turing** (1912-1954). Nel 1936 Turing pubblica l'articolo *On computable numbers*, nel quale immagina un dispositivo che può assumere una varietà di configurazioni, dette «**stati**», e che dispone di una testina per accedere a un nastro scorrevole contenente una sequenza di «**simboli**». A ogni istante il dispositivo si trova in un certo stato *X* e

LESSICO FILOSOFICO

funzionalismo è la concezione della mente secondo cui uno stato mentale non è identificabile con uno stato fisico del cervello, ma con la *funzione* che svolge nell'elaborazione delle informazioni.

intelligenza artificiale è l'ambito di ricerca interdisciplinare che punta far sì che una macchina dotata di regole per la combinazione di simboli – cioè un computer – riesca a compiere operazioni simili a quelle compiute dalla mente umana.

la sua testina punta su un certo simbolo Y . Il dispositivo è governato da un programma che stabilisce, per ogni possibile coppia X - Y , quale sarà il nuovo stato da assumere e quale l'operazione da compiere sul nastro (tre le opzioni: spostarlo a destra, spostarlo a sinistra, scrivere un simbolo nella casella su cui punta la testina).

Questo semplice dispositivo, che prende il nome di “macchina” di Turing, è la più potente **macchina di calcolo** che si possa concepire; la sua ideazione prefigura la costruzione dei primi computer, che avrà luogo nel decennio successivo.

Il test di Turing Nel 1950, nell'articolo *Computing machinery and intelligence*, Turing propone un test per stabilire se a un certo programma informatico si possa attribuire un'intelligenza paragonabile a quella di una persona. Immaginiamo che un esaminatore dialoghi con due soggetti, uno umano e uno artificiale, senza poterli vedere. L'idea è che se il soggetto artificiale riesce a partecipare al dialogo **ingannando l'esaminatore** con la stessa efficacia con cui lo farebbe una persona (ad esempio un uomo che cercasse di farsi passare per una donna), allora lo si può considerare **intelligente** come una persona.

Se, con la sua “macchina”, Turing aveva posto le condizioni per il nascere dell'informatica e dell'intelligenza artificiale, con il suo “test” indica l'obiettivo forse più ambizioso che questi ambiti di ricerca possano perseguire.

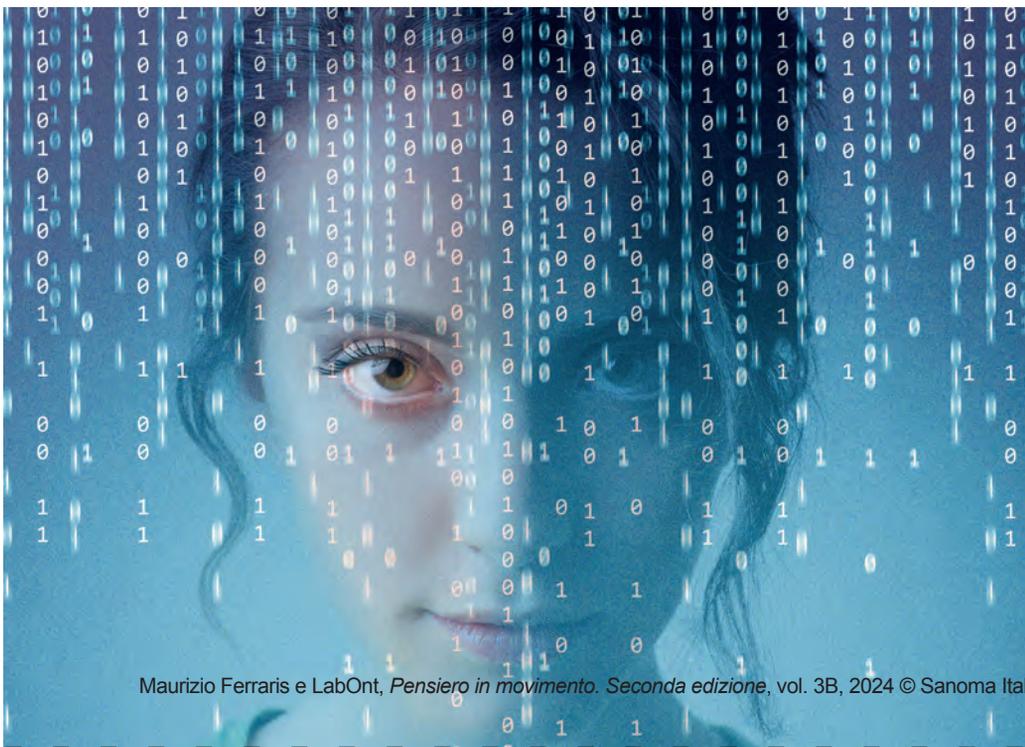
■ **Searle e la stanza cinese**

L'analogia fra mente e computer su cui si basano le scienze cognitive riscuote grande successo nella filosofia del tardo Novecento, ma riceve anche importanti critiche. In particolare, nel 1980 il filosofo americano **John Searle** (nato nel 1932) pubblica l'articolo *Mind, brains and programs*, in cui propone un esperimento mentale denominato “**stanza cinese**”. L'obiettivo di Searle è mostrare che, per attribuire intelligenza a un computer, non basta che questo superi un “test” come quello suggerito da Turing, risultando apparentemente indistinguibile da una persona.



PODCAST

Intelligenza artificiale
o intelligenza
naturale?
(Platone, Searle)



I termini dell'esperimento mentale Searle immagina che John (ovvero lui stesso), un madrelingua inglese che ignora il cinese, sia chiuso in una stanza e riceva dall'esterno sequenze di ideogrammi cinesi che costituiscono domande. Comportandosi come farebbe un computer, John li combina in base alle regole di un manuale, e consegna all'esterno della stanza nuove sequenze di ideogrammi che costituiscono le risposte a quelle domande. Dall'esterno si ha dunque l'impressione che John *parli* il cinese, ma è evidente che egli non *comprende* il cinese nel modo in cui comprende invece l'inglese.

Il richiamo all'intenzionalità della mente La conseguenza tratta da Searle è che il comprendere una lingua – e più in generale il **pensare** – non sia riducibile a una **mera elaborazione di dati**. Questo perché il pensare comporta anche un essenziale rivolgersi del soggetto a un oggetto, cioè comporta quella caratteristica essenziale della vita mentale che i filosofi chiamano **"intenzionalità"**. Quando John comprende il significato della parola inglese *flower*, la sua mente si rivolge a un fiore, mentre quando all'interno della stanza egli elabora l'ideogramma che rappresenta un fiore, la sua mente non si rivolge a nulla al di là dell'ideogramma medesimo. Ecco perché **l'intelligenza umana**, secondo Searle, è essenzialmente **diversa dall'intelligenza artificiale**.

■ Il paradigma della "cognizione situata"

Sul finire del Novecento, l'analogia fra la mente e il computer viene messa in discussione all'interno delle stesse scienze cognitive, a partire dall'osservazione che la **mente** è **legata al corpo**, e di conseguenza **al proprio ambiente**, in un modo molto più stretto e significativo di quello in cui un computer è legato al proprio ambiente. Se si vuole paragonare la cognizione umana a quella artificiale, allora il termine di paragone più adeguato non è il computer, bensì il **robot**, poiché quest'ultimo è in grado di agire nell'ambiente con modalità analoghe a quelle di un corpo vivente.

Le quattro E L'attenzione per il ruolo del corpo e dell'ambiente nello studio della cognizione umana ha portato in anni recenti all'affermarsi di un nuovo paradigma, che viene chiamato **cognizione situata**. L'espressione rimanda all'idea secondo cui la mente va studiata tenendo conto della **situazione concreta** in cui l'organismo opera. Si parla anche di paradigma delle **"quattro E"**, perché tale situazione è rappresentabile mediante quattro caratteristiche che in inglese sono tutte designate da parole che iniziano con la "E": *embodied* (incarnata), *embedded* (incastonata), *enactive* (enattiva) e *extended* (estesa). Ogni processo cognitivo è "incarnato" nel senso che dipende significativamente dal **corpo** nel suo complesso, e non soltanto dal cervello; è "incastonato" nel senso che è fortemente legato all'**ambiente naturale e sociale** in cui l'organismo si trova; è "enattivo" (dall'inglese *to enact*) nel senso che comporta un'azione dell'organismo nel suo ambiente; è "esteso" nel senso che le dinamiche cognitive si estendono **al di là dei confini corporei dell'organismo**, includendo funzioni svolte da apparati tecnici.

LESSICO FILOSOFICO

cognizione situata è l'espressione con cui si suggerisce che l'acquisizione e l'elaborazione dell'informazione da parte della mente umana non avvengono in un ambiente asettico (simile a un

laboratorio scientifico), ma in una situazione determinata, cioè attraverso un corpo collocato in un contesto naturale e sociale, con cui l'organismo interagisce e nel quale opera.

L'esempio del taccuino Per chiarire quest'ultimo punto, in un articolo del 1998 intitolato *La mente estesa*, Andy Clark (nato nel 1957) e David Chalmers (nato nel 1966; ►p. 527) fanno l'esempio del taccuino, il quale **estende la memoria** di un malato di Alzheimer permettendogli di *ricordare* ciò che altrimenti dimenticherebbe.

■ **Le reti neurali e il connessionismo**

Dagli anni Cinquanta agli anni Ottanta del Novecento, l'intelligenza artificiale si è basata sull'analogia tra i programmi informatici e i nostri processi di pensiero: processi dei quali si può fare esperienza diretta. In breve, gli informatici cercavano di **replicare** mediante i computer il **modo di pensare umano**.

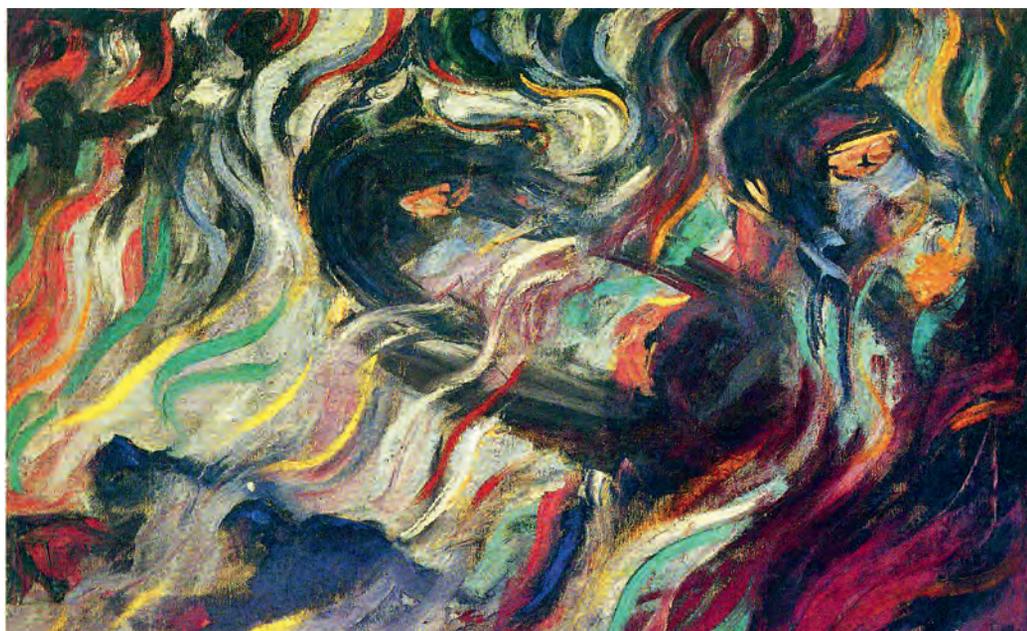
Dalla mente al cervello A partire dalla seconda metà degli anni Ottanta, gli informatici hanno invece cominciato a cercare di **replicare** il **modo di funzionare del cervello**, il quale, pur essendo lo strumento del nostro pensare, elabora le informazioni in modo assai differente rispetto a quanto noi riusciamo a comprendere e descrivere riflettendo sui nostri processi di conoscenza e di ragionamento. La mente consiste infatti di stati mentali che compiono operazioni cognitive su contenuti complessi, muniti di significato; il cervello invece è un sistema di neuroni (le cellule del sistema nervoso) che compiono operazioni su contenuti semplici, che di per sé non hanno alcun significato; il significato emerge piuttosto da un'enorme quantità di neuroni che si inviano segnali, creando in tal modo una "rete" costituita da nodi e connessioni tra nodi.

I sistemi di intelligenza artificiale progettati in modo da simulare il funzionamento del cervello prendono il nome di **reti neurali artificiali**.

rete neurale artificiale è il nome dato a un sistema artificiale di elaborazione delle informazioni che simula quello del sistema nervoso umano,

il quale si basa su migliaia di neuroni e sulle complesse "reti" di segnali che essi si scambiano.

**LESSICO
FILOSOFICO**



Umberto Boccioni,
*Stati d'animo:
gli addii*, 1911,
olio su tela.

Il modello connessionista L'affermazione delle reti neurali nell'intelligenza artificiale è avvenuta parallelamente all'affermazione, nell'ambito delle scienze cognitive, del paradigma delle "quattro E". Si è così assistito a uno spostamento di interesse dalla mente intesa in senso astratto alla concretezza del cervello. Poiché il cervello consiste di neuroni *connessi* tra loro, questa nuova tendenza che accomuna le scienze cognitive e l'intelligenza artificiale è stata definita **connessionismo**.

■ Il *machine learning* e l'intelligenza artificiale generativa

Sistemi che apprendono Nell'ambito dell'intelligenza artificiale, lo sviluppo del connessionismo e delle reti neurali, unito alla pervasività di Internet, ha portato alla creazione di nuovi potentissimi sistemi informatici capaci di **apprendere** nel modo in cui apprende il nostro cervello. Come il cervello umano impara continuamente, beneficiando dell'educazione familiare o scolastica e delle proprie esperienze, così i nuovi sistemi di intelligenza artificiale apprendono dall'enorme quantità di dati reperibile in Internet, migliorando progressivamente le proprie prestazioni. Si parla a tal proposito di *machine learning*, per indicare una forma di apprendimento automatico.

Sistemi che generano contenuti Dal momento che può rispondere alle richieste degli utenti "**generando**" nuovi contenuti (testi, immagini, audio o video), l'intelligenza artificiale che sta alla base di questi sistemi è chiamata "**intelligenza artificiale generativa**". Il più celebre di tali sistemi è attualmente **ChatGPT** (acronimo di *Chat Generative Pre-trained Transformer*), un software creato nel 2022 nel laboratorio OpenAI. ChatGPT è in grado di dialogare con un utente umano su qualsiasi argomento, comportandosi come

farebbe una persona, ma con una maggiore rapidità di ragionamento e di risposta, e basandosi su una mole di conoscenze pressoché inarrivabile per gli esseri umani.

Dialogando con ChatGPT si ha l'impressione di interagire con un soggetto cosciente come noi. Ma è davvero così? Per capirlo, occorre soffermarsi sulla nozione stessa di coscienza.



LESSICO FILOSOFICO

connessionismo è un paradigma teorico che studia i processi cognitivi mediante le reti neurali, e che porta a programmare i computer ricreando

artificialmente la struttura di connessioni che si instaura tra le cellule del nostro sistema nervoso.