

Macchine che pensano
Un itinerario nell'Intelligenza Artificiale, tra letteratura,
informatica, matematica e filosofia

a cura degli studenti **Alfredo Cesarano, Ciro Strazzullo (3T),**
Francesco Bava, Giuseppe Castrale, Giovanni Iengo,
Fabio Vuolo (4T), Christian Coscia, Matteo Diaco (5T)
dell'ITI "G. Ferraris" di Napoli
coordinati dai docenti **Francesca Carleo, Clara Di Antonio,**
Maria Di Maro, Giuseppe Mangione,
Angelo Rosano, Ciro Totaro

L'articolo sintetizza gli interventi presentati al Terzo Congresso ANFSU "Intelligenza umana e intelligenza artificiale", tenutosi il 9 maggio 2025 presso l'ISIS "E. Caruso" di Napoli, ed al Convegno "Studenti in cattedra, docenti nei banchi", svoltosi il 24 maggio 2025 presso il Liceo Scientifico "Don Milani" di Gragnano. Il lavoro è il frutto di un'attività interdisciplinare che ha coinvolto l'intero triennio della sezione T del "Ferraris", in sinergia con il progetto "Filosofando", che da anni propone, in questo Istituto, l'introduzione di tematiche filosofiche ed artistiche. A partire dalle suggestioni letterarie di Isaac Asimov e dalle riflessioni di Alan Turing, si sono affrontate alcune delle questioni più dibattute, da studiosi di diverse discipline, sulla natura dell'IA e sulle implicazioni etico-filosofiche del suo straordinario sviluppo. Si è inoltre evidenziato come molti temi al centro del dibattito contemporaneo, e molte posizioni teoriche relative ad esso, abbiano avuto autorevoli precedenti nel pensiero dei massimi filosofi del Seicento e del Settecento.

Asimov visionario, ovvero a che servono i docenti, ci bastano le macchine

Nel 1976, **Isaac Asimov**, geniale scrittore di fantascienza, di nazionalità statunitense e di origine ebraico-russa, pubblicò un breve racconto intitolato “I nuovi docenti”, dove immaginava un futuro in cui i robot avessero sostituito gli insegnanti, all’interno di un “sistema educativo perenne” che doveva svolgersi per tutta la vita. Il racconto dà indicazioni molto precise: immagina “macchine istruttrici” complesse e flessibili, in grado di modificare il proprio programma sulla base degli input ricevuti dagli studenti. Lo scrittore descrive nel dettaglio anche le modalità dell’interazione: *“Lo studente farà domande, risponderà a domande, farà asserzioni, offrirà opinioni, e da tutto questo la macchina sarà in grado di valutarlo in modo tale da poter regolare la velocità e l’intensità del proprio corso d’insegnamento e andare incontro agli interessi mostrati”* (Asimov, 2024, p. 415).

Ma Asimov non si ferma qui: delinea una realtà futura con biblioteche centrali completamente computerizzate e interconnesse, tanto da formare una “singola biblioteca planetaria”, alla quale tutte le “macchine istruttrici” verrebbero collegate, ed in cui sarebbe possibile avere a disposizione qualsiasi testo, sia in forma di libro o documento, sia registrato come file audio o video. Ma gli “insegnanti umani” che fine farebbero? Asimov concede che “non verrebbero completamente eliminati”. In alcune materie dove “l’interazione umana è essenziale”, come “la ginnastica o la recitazione”, potrebbero continuare a svolgere il loro compito, ridotto però

a quello di semplici facilitatori (i nostri docenti ci hanno assicurato che tra le idee dei “riformatori” della scuola, circolano pensieri non molto lontani da questi).

Comunque Asimov si mostra sensibile al tema della relazione umana nel processo di apprendimento; infatti elogia il lavoro di gruppo tra “studenti operanti in un campo specifico” e sottolinea anche l’importanza delle riunioni tra studenti ed “esperti umani”, in cui si formulano ipotesi e si sviluppano intuizioni, che però vanno avvalorate o confutate dal “ritorno” alle “loro macchine interminabilmente informate” (Asimov, 2024, p. 416). All’obiezione che insegnare presuppone un pensiero, un’intelligenza, e invece i computer-robot possono fare solo ciò per cui sono stati addestrati, Asimov risponderà, in un successivo racconto pubblicato nel 1977, che questo è vero, ma “anche il cervello può fare solo ciò per cui è stato programmato dai suoi geni” (Asimov, 2024, p. 425).

Già in questo breve cenno ai racconti di Asimov possiamo trovare le domande essenziali che ci siamo posti in questo lavoro: ***si può affermare che i computer “pensino”? Che differenza c’è tra intelligenza artificiale ed umana?***

La risposta di Alan Turing

In un saggio del 1950 Alan Turing sostiene che **è possibile che una macchina possa pensare**. Immagina che, potenziando e perfezionando i calcolatori elettronici, si potessero ottenere macchine con prestazioni difficilmente distinguibili da quelle umane, sia per quanto riguarda la comprensione e la produzione del linguaggio, sia in riferimento alla capacità di risolvere problemi (Turing, 2025).

Nello stesso saggio Turing risponde alle obiezioni più forti rispetto alla tesi di macchine pensanti; vediamone le principali:

1) l'obiezione della **coscienza**: pensare, infatti, non significa solo comporre concetti, ma anche avere la coscienza di farlo; 2) l'obiezione **logico-matematica**: il teorema di Gödel dimostrerebbe che in qualsiasi sistema di logica si possono formulare asserzioni che non possono essere né dimostrate né confutate all'interno del sistema stesso, a meno che questo non sia contraddittorio; 3) l'obiezione della **creatività**: la macchina sa fare soltanto ciò che le abbiamo insegnato a fare, invece pensare significa avere capacità creative.

Turing supera la prima obiezione con una considerazione di senso comune: noi attribuiamo il pensiero ai nostri simili basandoci sul loro comportamento, certamente non accedendo alla loro consapevolezza. Anche la seconda obiezione è facilmente superabile con considerazioni di questo tipo: attribuire dei limiti al pensiero logico-matematico non vuol dire dimostrare che tali limiti non esistano anche nel pensiero umano. La terza obiezione è più difficile da confutare: il pensiero non è solo capacità di applicare la logica deduttiva, ma è anche fare inferenze, induzioni, costruire “concetti nuovi”. In realtà Turing, in questo saggio, risponde all'obiezione limitandosi a proporre metodologie capaci di sviluppare nelle macchine nuove capacità e prestazioni, che somigliano all'“addestramento” degli attuali sistemi di intelligenza artificiale generativa. Nel saggio, infine, c'è una prima versione del famoso “test di Turing”, che lo stesso autore perfezionerà in altre occasioni, e che si basa sull'assunto che pensiero umano e artificiale sarebbero assimilabili nel caso in cui un interrogante non

riuscisse, in un'apprezzabile percentuale di prove, a distinguere il comportamento verbale di un computer da quello di un essere umano (il test ha avuto parecchie rielaborazioni e, di recente, è stato ampiamente superato da modelli linguistici come GPT-4.5, scambiati per esseri umani con percentuali superiori al 70%).

Intelligenza artificiale e pensiero umano: posizioni a confronto

Dall'anno di pubblicazione del saggio di Turing ad oggi, il dibattito sulla possibilità di ricostruire artificialmente i meccanismi dell'intelligenza umana ha coinvolto studiosi di diverse discipline: filosofia, neuroscienze, informatica, matematica, psicologia.

Sintetizzando estremamente, possiamo raggruppare le varie posizioni in due grandi maxi-insiemi: coloro che sostanzialmente assimilano pensiero umano e intelligenza artificiale e quelli che invece ritengono improponibile tale assimilazione. All'interno di questi due grossi schieramenti, troviamo però posizioni molto diversificate su questioni importanti, riguardanti, ad esempio, il significato della coscienza e del libero arbitrio o la dimensione degli eventuali pericoli di uno sviluppo incontrollato dei sistemi di intelligenza artificiale.

Nel primo gruppo possiamo includere i pionieri dell'IA che seguono la scia di Turing, come i matematici **John McCarty** e **Marvin Minsky**: il primo nel 1956 coniò il termine “intelligenza artificiale” e definì l'intelligenza “la parte computazionale della capacità di raggiungere obiettivi”, il secondo, riprendendo un'idea che era stata già introdotta dal filosofo David Hume, intese l'intelligenza artificiale come il prodotto di una miriade di “agenti subpersonali” che si organizzano spontaneamente. Nello stesso gruppo

vanno inseriti i teorici della mente cosiddetti “funzionalisti” (filosofi, logici e informatici) per i quali la coscienza e il pensiero sono forme di elaborazione di informazioni che non necessitano di cervelli biologici per essere implementate.

In questo primo “insieme” vi poi sono studiosi che superano l’obiezione della “coscienza” proponendo una concezione rigorosamente materialistica della stessa. Tra questi citiamo il neuroscienziato **Anil Seth**, per il quale la coscienza è un processo biologico legato al semplice essere vivi: *essere sé* riguarda quindi essenzialmente *il corpo*. Questo studioso, infatti, identifica i possibili futuri rischi non tanto nello sviluppo di macchine intelligenti, quanto nella futuribile interazione tra queste macchine e lo sviluppo di tecniche di *editing genetico* che prospettano la possibilità di creazione di “nuove forme di vita”, come, ad esempio, gli “organoidi cerebrali”, piccole strutture derivate da cellule staminali che “potrebbero ospitare forme primitive di consapevolezza corporea” (Seth, 2024, pp.275-276). Inoltre citiamo il filosofo **Daniel Dennett**, per il quale la coscienza non fa capo ad un centro che possiamo chiamare *io*, ma emerge dalla collaborazione e dalla competizione di parti del cervello autonome; sostiene infatti Dennett, parafrasando Minsky e Hume, che gli esseri umani sono come “robot fatti di robot” di dimensioni sempre più piccole.

Nell’altro gruppo, invece, possiamo comprendere coloro che non ritengono possibile assimilare intelligenza umana ed artificiale. Innanzitutto qui ritroviamo gli studiosi che ribadiscono l’unità profonda di corpo e mente, come **Antonio Damasio**, neuroscienziato che attribuisce ad emozioni e passioni un ruolo chiave nei processi cognitivi, ma che non esclude la possibilità di macchine “in grado di sentire” attraverso lo sviluppo della

robotica, per giungere ad “ibridi di creature naturali e artificiali, efficaci assistenti di esseri umani realmente senzienti” (Damasio, 2022, p.184). Ci sono poi gli studiosi che negano che i sistemi di IA abbiano capacità semantiche, come il filosofo **John Searle**, per il quale questi sistemi non sono in grado di attribuire significati a simboli, ma operano solo associando e traducendo segni in altri segni.

Tra gli studiosi italiani bisogna menzionare il filosofo **Luciano Floridi**, che è scettico sullo sviluppo delle capacità “intelligenti” dell’IA e sostiene, in modo ironico ma efficace, che nell’IA non ci sia più intelligenza di “quella presente in un tostapane”, aggiungendo che la fortuna dell’IA derivi non tanto dalla sua effettiva “intelligenza”, ma dal fatto che stiamo costruendo un ambiente sempre più ricettivo all’utilizzo di questo tipo di “intelligenza”, estremamente utile, ma sostanzialmente “stupida”, ed aggiunge che dovremmo preoccuparci “della vera stupidità umana, non dell’intelligenza artificiale immaginaria” (Floridi, 2022, p. 276). È chiaro che questa posizione si basa su una nozione di intelligenza umana che non può essere limitata alla sola capacità di calcolo.

Intelligenza e calcolo: anticipazioni di alcune tematiche del dibattito contemporaneo sull’IA nei filosofi del Seicento e del Settecento

Come abbiamo visto nel paragrafo precedente il dibattito contemporaneo sulla natura dell’IA ruota attorno alla questione fondamentale della definizione stessa di “intelligenza”, ed è sorprendente come molte anticipazioni di questi temi le ritroviamo nelle riflessioni dei più importanti filosofi del Seicento e del Settecento.

Il primo riferimento fondamentale è l'opera di **Renato Cartesio**, grande matematico e filosofo, sostenitore di un deciso **dualismo mente-corpo**. Dice Cartesio che posso dubitare di tutto, ma, mentre dubito, non posso dubitare che sto pensando: da cui il famoso motto “cogito ergo sum”, cioè *penso dunque sono*. L'uomo è, fondamentalmente, “una cosa che pensa”, una “sostanza pensante”. Mente e corpo sono dunque indipendenti: tutto l'universo è costituito da due sostanze distinte, materia (*res extensa*) e spirito (*res cogitans*), ed è retto da principi differenti. Per Cartesio, quindi, non esiste una correlazione diretta tra stati mentali e stati corporei e l'uomo si caratterizza per l'*autocoscienza* (capacità di pensarsi come *cosa che pensa*) e per una razionalità separata dalle passioni. Il dualismo cartesiano è ripreso oggi dai fautori dell'IA “intelligente”, per i quali la razionalità non è legata alle emozioni, ed è invece avversato dai sostenitori dell'unione mente-corpo, che negano l'assimilazione dell'IA all'intelligenza umana.

Un altro pensatore fondamentale è il filosofo inglese **Thomas Hobbes**, il quale è tra i primi a definire il pensiero come “computazione”: **pensare**, per Hobbes, **non è altro che calcolare**, sviluppare in modo logicamente coerente schemi matematici di ragionamento (*algoritmi*). Questa visione radicale definisce la mente umana come “macchina” calcolatrice, e riduce la complessità del pensiero umano a semplici operazioni aritmetiche, assimilando anche il linguaggio allo sviluppo logico-matematico di algoritmi. Nel fare questo, Hobbes precorre l'avvento dell'IA e le tesi che affermano l'identità tra la produzione del pensiero e il funzionamento materiale del cervello.

Sulla stessa linea di Hobbes troviamo **Gottfried Leibniz**, filosofo, logico e matematico tedesco, che collegò *ars combinatoria* e calcolo e pose le basi

per costruire una “lingua artificiale universale”, in grado di risolvere i problemi logici così come si risolvono i problemi algebrici. In un testo del 1666 descrisse il “sogno” che si augurava: “quando sorgeranno controversie fra filosofi, non sarà più necessaria una discussione, come non lo è fra calcolatori. Sarà sufficiente che essi si siedano di fronte agli abachi e si dicano l’un l’altro: **Calculemus!**”. Le sue tesi sono alla base del *funzionalismo*, che assimila la mente al programma di un calcolatore, il quale, tramite algoritmi, manipola simboli.

Un filosofo del Seicento che invece ribadisce la profonda **unità tra mente ed emozioni** è **Baruch Spinoza**, che il neuroscienziato Damasio, non a caso, assume a modello di riferimento della sua concezione non “computazionale” dell’intelligenza umana.

Per quanto riguarda la natura della *mente pensante*, molto rilevante è la posizione di **David Hume**, che descrive l’io come un palcoscenico in cui si presentano volontà e istanze anche contraddittorie, e ipotizza una “**mente senza un sé**”: l’io è ridotto ad un “fascio di impressioni” che si susseguono nel tempo. Come già visto, si tratta delle tesi riprese da Minsky e Dennett, che scompongono le prestazioni cognitive in sottosistemi dotati di autonomia: come scriverà Minsky, il pensiero non sarebbe il risultato di una specifica funzione che possiamo chiamare io, ma di una sorta di società *della mente*.

Chiudiamo con **Immanuel Kant**, che è a nostro avviso il pensatore che fornisce gli strumenti più utili per differenziare intelligenza umana e artificiale. Kant individua tre diverse dimensioni costitutive della mente, di cui quella conoscitiva in senso stretto è l’**intelletto**, che permette di

formulare giudizi scientifici basati su “condizioni a priori” comuni a tutti gli uomini (in via ipotetica tali condizioni potrebbero essere trasferite come software di base in sistemi di IA). Ma, accanto all’*intelletto*, Kant individua altre due dimensioni, la **ragione**, i cui principi ispirano la vita morale e la **facoltà di giudizio**, che riflette il sentimento di appartenenza al genere umano nella sua universalità, disegnando una **complessità della mente e del pensiero** umani che non possiamo assimilare in nessuna misura a sistemi di Intelligenza Artificiale.

Conclusioni: etica ed Intelligenza Artificiale

A partire dalla visione kantiana della “complessità umana”, riprendiamo le tesi di Floridi, il quale individua il vero problema dell’IA non tanto in fantascientifici sviluppi verso forme di superintelligenza artificiale, ma nella delega acritica, a questi sistemi, di comportamenti e scelte che implicano problematiche relazionali ed etiche.

In classe abbiamo analizzato un caso accaduto nel 2023, in Belgio, dove un uomo di trent’anni, sposato e padre di due bambini, fu indotto al suicidio da una frequentazione intensa con un *chatbot*, progettato e sviluppato da una società californiana, che offriva a chiunque la possibilità di crearsi un amico virtuale con il quale stabilire qualsiasi genere di relazione e conversare su ogni argomento.

Come dice Guido Scorza, componente del Collegio del Garante per la protezione dei dati personali, “Se non si prova, sembra una cosa da folli. Ma non lo è. È un enorme mercato di sentimenti, emozioni, affetti sintetici basato sull’IA, straordinariamente florido” (Scorza, 2024). Un mercato

fondato sulle debolezze e la fragilità umane, alimentate dalle chiusure al mondo reale sempre più diffuse nella nostra società.

Possiamo quindi concludere, parafrasando Kant, che l'IA dovrebbe essere usata per trattare le persone ***sempre come fini in sé e mai come mezzi***, perché gli uomini, a differenza delle macchine, sono creature dotate di una **volontà** che, liberamente, **stabilisce dei fini** e deve sforzarsi di realizzarli senza intaccare la libertà degli altri.

* In classe sono stati letti e commentati brani tratti dalle seguenti opere:

Asimov I. (2024). *Visioni di robot*. Milano: Feltrinelli.

Cartesio R. (1996). *Discorso sul metodo*. Milano: RCS Libri.

Damasio A. (2022). *Sentire e conoscere*. Milano: Adelphi.

Floridi L. (2022). *Etica dell'intelligenza artificiale*. Milano: Cortina.

Kant I. (2004). *Critica della capacità di giudizio*. Milano: Rizzoli.

Searle J. (2023). *Intelligenza artificiale e pensiero umano*. Roma: Castelvecchi.

Seth A. (2023). *Come il cervello crea la nostra coscienza*. Milano: Cortina.

Scorza G. (2024). *Istruzioni per morire dal mio amore artificiale*, <https://www.garanteprivacy.it>.

Turing A. (2025). *Macchine calcolatrici e intelligenza*. Torino: Einaudi.