Agli albori dell'AI. Riflessioni sul test di Turing

di Francesco Federico Calemi

Over seventy years after its publication, Alan Turing's article Computing Machinery and Intelligence still fuels debate on the meaning and structure of the test it introduces, the limits of answering "Can machines think?", and the validity of that question itself. Widely seen as the first visionary manifesto on artificial intelligence, the article anticipated many ideas that became foundational in AI research. At a time when computers were mainly tools for high-speed calculation – first in military contexts and then in science – Turing used the concept of a computational machine to pose philosophical questions about human intelligence, the link between computation and thought, and the possibility of algorithmically modeling cognition. This essay reconstructs the test's structure and assumptions (§§ 1-3), explores the link between Turing's and Leibniz's views on thinking machines (§ 4), and examines major criticisms of the mental model in Turing's work (§ 5).

Keywords: Turing Machine; Functionalism; Imitation Game; Wittgenstein; Gödel

L'intuizione di Turing

L'elaborazione del «gioco dell'imitazione» — noto come *test di Turing*, in omaggio al suo ideatore — avvenne in un periodo storico vicino alla Seconda Guerra Mondiale e ancora profondamente segnato dai suoi tragici avvenimenti. Proprio tra il 1940 e la conclusione del conflitto, Alan Turing, lavorando segretamente nell'unità di controspionaggio britannico riunita a Bletchley Park, contribuì in modo determinante allo sviluppo di *The Bombe*, un dispositivo concepito per decifrare i codici nazisti generati dalla macchina *Enigma*[1]. Come scrive Copeland:

> La guerra fu una parentesi disastrosa per molti, ma per Turing rappresentò uno sbocco di importanza internazionale per il suo genio creativo. Non è esagerato affermare che, senza Turing, la guerra sarebbe probabilmente durata più a lungo e avrebbe potuto essere vinta dai nazisti[2].

Infatti, l'uso dei calcolatori in ambito bellico rivelò l'enorme potenziale rappresentato dalle macchine nel gestire compiti complessi e stratificati, come il ragionamento logico e la ricerca euristica, attività tipicamente associate all'intelligenza umana. Fu proprio nel corso di questo conflitto devastante che la ricerca scientifico-tecnologica compì progressi straordinari nell'ambito della computazione. In particolare, secondo una testimonianza raccolta da Jack Copeland in un'intervista a Donald Michie, collega di Turing a Bletchley Park, sembra che già in quegli anni Turing stesse maturando un'intuizione di vasta portata[3]: che i calcolatori non solo potessero essere programmati per risolvere problemi tecnici specifici, ma che, in linea di principio, avrebbero potuto simulare processi cognitivi umani, gettando le basi per quella che sarebbe poi divenuta la moderna intelligenza artificiale (AI), rispetto alla quale il test di Turing rappresenta un passo fondamentale e preliminare. Cercheremo, in quanto segue, di chiarire in cosa consista il test e quale sia il suo significato logico-filosofico.

---

1 - Rinviamo per maggiori approfondimenti alla raccolta J. Copeland *et al.* (a cura di), *The Turing Guide*, Oxford University Press, Oxford 2017.

2 - *Ibid.*, p. ix.

3 - «After a heavy week of codebreaking, the two would meet in a village pub on Friday evenings to discuss how to reproduce human thought processes in a universal Turing machine […] From chess Turing moved on to the idea of automating the whole process of learning» (J. Copeland, *Intelligent Machinery*, in J. Copeland *et al.* [a cura di], *The Turing Guide*, cit., p. 267).

Il gioco dell'imitazione

Come è noto, Turing dedica la prima sezione dell'articolo *Computing Machinery and Intelligence*[4] alla formulazione (e alla successiva ri-formulazione) di quella che è la domanda-guida dell'intero scritto: «le macchine possono pensare?». La provocatorietà di tale interrogativo potrebbe suggerire una lettura semplificata e ingenua secondo la quale il test mostrerebbe che le macchine, in qualche modo, possano pensare esattamente nello stesso senso in cui diciamo che gli esseri umani pensano. Senza attribuire a questa lettura un effettivo seguito in ambito filosofico, la considereremo come una posizione limite, utile a mettere in luce – per contrasto – il corretto quadro teorico entro cui si può cogliere il significato autentico del test di Turing.

Anzitutto lo stesso Autore precisa che per rispondere in modo diretto al quesito «le macchine possono pensare?» occorre un preliminare chiarimento dei termini "macchina" e "pensare". Tuttavia, osserva Turing, ciò non può avvenire imponendo delle definizioni, né cercando di comprendere il modo in cui adoperiamo tali espressioni nelle nostre pratiche linguistiche quotidiane. Nel primo caso si rischierebbe, infatti, di predeterminare arbitrariamente una risposta, mentre nel secondo caso il tentativo di fornire una risposta richiederebbe, assurdamente, una indagine demoscopica. Turing, a fronte di queste difficoltà, propone una peculiare riformulazione della domanda iniziale per mezzo dell'espediente del «gioco dell'imitazione»[5], di cui è opportuno esporre sinteticamente set up, scopo e svolgimento.

Il gioco dell'imitazione prevede tre giocatori, un uomo (giocatore A), una donna (giocatore B) e un interrogante, situati in stanze differenti e reciprocamente isolate in modo che nessuno dei tre possa avere accesso alle altre stanze. (Questa condizione di inaccessibilità fisica risulterà, come vedremo, di estrema importanza per comprendere il problema dell'attribuzione di stati mentali.) L'interrogante può tuttavia comunicare con l'uomo e la donna tramite una telescrivente, mentre non è previsto che l'uomo e la donna comunichino in alcun modo tra di loro, né che in qualche modo possano dare prova effettiva di quanto di volta in volta dichiarano.

Le condizioni di vittoria del gioco sono le seguenti: lo scopo dell'interrogante è determinare chi è l'uomo e chi è la donna solo per mezzo dell'intercorsa comunicazione scritta. Lo scopo del giocatore B è quello di aiutare l'interrogante ad effettuare una corretta identificazione, mentre lo scopo del giocatore A è quello di depistare l'interrogante impedendogli di effettuare una corretta identificazione.

---

4 - A. Turing, *Computing Machinery and Intelligence*, in «Mind», 59 (1950), pp. 433-460. Nel presente lavoro faremo riferimento alla seguente traduzione italiana: A. Turing. *Meccanismo computazionale e intelligenza artificiale*, tr. it. a cura di L. Cimmino, Città Nuova, Roma 2018.

5 - Turing, *Meccanismo computazionale e intelligenza artificiale*, cit., pp. 7-8.

Per dirlo con le parole di Turing, per B «la migliore strategia potrebbe essere quella di fornire risposte vere»[6] laddove la migliore strategia di A è quella di far sì che l'interrogante «sbagli l'identificazione»[7].

Naturalmente per l'interrogante risulta altamente difficoltoso vincere al gioco dell'imitazione dato che i soli dati di cui dispone per poter effettuare le proprie identificazioni sono di tipo linguistico: nessuna ispezione visiva, tattile o comunque sensoriale gli è permessa nel tentativo di indicare, tra i due giocatori coi quali comunica con telescrivente, chi è l'uomo e chi è la donna.

Definito il gioco dell'imitazione, Turing chiede quanto segue:

> "Cosa accade se è una macchina ad assumere il posto di A nel gioco?". L'interrogante sbaglierà tanto spesso quanto avverrebbe se a giocare fossero un uomo e una donna?[8].

Turing aggiunge: «Queste domande sostituiscono la domanda iniziale "Le macchine possono pensare?"»[9]. In altri termini, la domanda-guida dello scritto di Turing diviene la seguente: un essere umano può ingannarsi nell'identificare e riconoscere, a partire dalla sola ispezione del comportamento linguistico, una macchina e un essere umano? Questa riformulazione rende chiaro che il focus della domanda non è il (presunto) soggetto-macchina e la sua (presunta) capacità di avere stati mentali, ma diviene propriamente il soggetto umano: quando e in quali condizioni attribuiamo l'attività del pensare? In quali condizioni giudichiamo *intelligente* un comportamento?

Naturalmente, nel quadro dell'esperimento mentale considerato, si dirà che una macchina "supera" il test di Turing se l'interrogante sbaglierà con sufficiente frequenza le sue identificazioni. Su questa possibilità Turing si esprime nel testo in senso pienamente positivo sostenendo quanto segue:

> Credo che tra cinquant'anni sarà possibile programmare i computer in grado di partecipare al gioco dell'imitazione così bene che un interrogante medio non avrà più del 70% di possibilità di operare la corretta identificazione dopo cinque minuti di domande[10].

Per quanto Turing si mostrasse fiducioso che tale soglia sarebbe stata raggiunta in tempi brevi, è solo nei primissimi anni Venti del XXI secolo che alcuni modelli linguistici avanzati – come GPT-4 – hanno cominciato a superare versioni rigorose del test, sia pure con circa venticinque anni di ritardo. È tuttavia importante preci-

---

6 - *Ibid.*, p. 8.

7 - *Ibidem.*

8 - *Ibidem.*

9 - *Ibidem.*

10 - *Ibid.*, pp. 19-20.

sare che si tratta di un traguardo solo parziale: tali sistemi, per quanto sofisticati nella simulazione del linguaggio umano, non possiedono ancora una vera intelligenza artificiale generale, e la loro "comprensione" resta confinata a domini specifici e a comportamenti appresi statisticamente. Tuttavia, ciò su cui focalizzeremo l'attenzione in questo contesto non è l'effettiva implementazione storica del test, ma la sua struttura concettuale e il tipo di attribuzioni mentali che esso implica.

La macchina universale di Turing

La seconda questione da chiarire per comprendere il significato del test di Turing è quella riguardante le macchine coinvolte nel gioco dell'imitazione. Le macchine a cui si fa riferimento – che qui denomineremo come di consueto *Macchine di Turing* (MT) – sono dispositivi progettati in stretta analogia con il modo in cui un essere umano svolge i propri calcoli, ossia in ottemperanza a regole operative specificate in un manuale o memorizzate, e svolgendo calcoli su fogli di carta (ipoteticamente disponibili in quantità illimitata). Per tale motivo una MT è composta da un'unità esecutiva, un'unità di controllo, e un dispositivo di archiviazione. In termini pratici, la macchina dispone di un nastro mobile, potenzialmente illimitato, suddiviso in riquadri su ciascuno dei quali è possibile stampare un solo simbolo alla volta, di un lettore ottico e un dispositivo di stampa. Ogni stato della macchina sarà pertanto determinato da una quintupla ordinata costituita da, rispettivamente, la configurazione iniziale della macchina in un dato momento, *t*, il simbolo scansionato in *t*, l'operazione da svolgere, il movimento del nastro, e la configurazione finale assunta dalla macchina[11].

Successivamente Turing generalizza la nozione di MT mettendo a punto il modello di una macchina universale, che in questa sede denomineremo *macchina di Turing universale* (MTU). Una MTU è in grado, infatti, di ricevere come input sul proprio dispositivo di archiviazione una rappresentazione aritmetizzata delle istruzioni operative di qualsiasi altra MT e di simularne perfettamente il comportamento[12]. Turing identifica quindi una MTU con un *computer digitale* ossia una macchina a stati-discreti tale da "saltare" da uno stato interno all'altro, e tale da rispondere a determinati input con adeguati output, imitando all'occorrenza, con un'apposita

11 - Per una descrizione più accurata si veda M. Davis. *Il calcolatore universale. Da Leibniz a Turing*, tr. it. A cura di G. Lolli, Einaudi, Torino 2012 e F.F. Calemi, *Macchine e pensiero. Assiomatizzazione e limiti della computabilità a partire da Turing*, in M. Falcone - E. Pili (a cura di), *L'irriducibile umano. Etica e diritto delle intelligenze artificiali*, Il Mulino, in corso di pubblicazione.

12 - Per un approfondimento rinviamo al saggio introduttivo di Gabriele Lolli in A. Turing, *Intelligenza meccanica*, tr. it. a cura di G. Lolli, Bollati Boringheri, Torino 1992.

programmazione, il comportamento di qualsiasi altra macchina a stati-discreti. Come rileva Copeland:

> leggendo ed eseguendo le istruzioni che il programmatore memorizza sul nastro, la macchina di Turing universale [...] è in grado di eseguire qualsiasi computazione possibile[13].

Turing è fermamente convinto che, poiché l'attività del pensare è legata al funzionamento del cervello umano e, in particolare, poiché il cervello umano può essere interpretato come una macchina a stati discreti (su questo punto cruciale dovremo tornare nel prosieguo), un computer digitale, inteso come una MTU, possa simularne il funzionamento. Di conseguenza, se il soggetto interrogante non riesce a superare il test di Turing, è ragionevole considerare "pensiero" anche il prodotto dell'attività di un computatore digitale, per quanto controintuivito ciò possa sembrare.

Per comprendere più a fondo tale prospettiva riteniamo utile proporre un parallelismo tra l'argomento di Turing e quello sviluppato da Leibniz nei *Principi della filosofia* in merito al problema delle "macchine pensanti": ciò aiuterà a comprendere l'innovatività della proposta di Turing in merito al dibattito sulla natura del mentale.

Dal mulino pensante al computatore digitale

In un celebre passo dei *Principi della filosofia* Leibniz invita il lettore ad immaginare, come parte di una *reductio ad absurdum* di ogni spiegazione meccanicistica del pensiero, di poter visitare e osservare dall'interno una presunta "macchina pensante", similmente a come si potrebbe entrare in un mulino osservandolo internamente. Con questa vivida immagine Leibniz sostiene che, anche scomponendo meccanicamente il cervello, non si sarebbe trovato nulla in grado di spiegare il pensiero, ma «solo parti che spingono altre parti»:

> D'altra parte si deve riconoscere che la percezione, e quel che ne dipende, *è inesplicabile mediante ragioni meccaniche*, cioè mediante le figure e i movimenti. Immaginiamo una macchina strutturata in modo tale che sia capace di pensare, di sentire, di avere percezioni; supponiamola ora ingrandita, con le stesse proporzioni, in modo che vi si possa entrare come in un mulino. Fatto ciò, visitando la macchina al suo interno, troveremo sempre e soltanto pezzi che si spingono a vicenda, ma nulla che sia in grado di spiegare una percezione. Quindi la [ragio-

13 - J. Copeland, *Turing's great invention: the universal computing machine*, in J. Copeland *et al.* (a cura di), *The Turing Guide*, cit., p. 54.

> ne della] percezione va cercata nella sostanza semplice, non già nel Composto, cioè nella macchina[14].

Ogni macchina, dunque, per quanto complessa, rimane un'entità puramente materiale, composta da parti meccaniche che si muovono secondo leggi fisiche prestabilite. Di conseguenza, nell'ipotesi che l'attività mentale sia semplicemente il prodotto di una macchina, e ipotizzando di poter estendere la scala della macchina fino a renderla esplorabile, troveremmo al suo interno solo movimenti e meccanismi, ma mai la *percezione*, la *coscienza*, o, in generale, il *pensiero*. Ne deriva che ogni tentativo di spiegare la mente esclusivamente in termini materialistici o meccanicistici risulta destinato al fallimento.

La posizione leibniziana potrebbe dunque essere riassunta dal seguente argomento: se la mente fosse una macchina, allora sarebbe una macchina pensante; se la mente fosse una macchina pensante, allora in essa vi sarebbero pensieri (percezioni, coscienza); ma in una macchina vi sono solo e soltanto parti meccaniche; dunque la mente non è una macchina. Non ci addentreremo nei dettagli dell'argomento di Leibniz[15], ma considereremo l'originalità dell'approccio al problema del mentale che emerge entro il contesto offerto da Turing e dal suo test.

Anzitutto occorre sottolineare che Turing legge l'analogia mente-macchina in una prospettiva funzionalista evitando il binomio materiale-immateriale: l'implementabilità della MTU mostra che il pensiero può essere descritto in termini di calcolo simbolico, rendendo la meccanicità del calcolo non una negazione del pensiero stesso, ma una sua rappresentazione in termini funzionali. Inoltre, e proprio per via di questa impostazione, Turing evita il problema della riducibilità degli stati mentali posto da Leibniz, mostrando che l'autentica questione non è il tentare di spiegare la natura del pensiero col "collocarlo" in uno spazio non-fisico (quello del mentale, per l'appunto), ma nella comprensione dei processi astratti che un supporto fisico può realizzare. Da ciò consegue che Turing, concentrandosi su ciò che una macchina può fare e su come il suo comportamento possa essere indistinguibile da quello umano, non cerca precipuamente di indagare la natura ontologica della mente, ma sposta l'attenzione sulla predisposizione umana ad attribuire (o negare) attività pensante in prossimità di comportamenti linguistici osservabili. Ciò consente di sostenere che il test di Turing ritrova il suo autentico scopo e significato nel quadro di una valutazione di quanto e come una macchina possa replicare comportamenti intelligenti *indipendentemente dalla loro* (*eventuale*) *natura intrin-*

---

14 - G. W. Leibniz, *Principi della filosofia o Monadologia*, trad. it. di S. Cariati, Bompiani, Milano 2001, p. 65.

15 - Per una discussione più approfondita rinviamo a P. van Inwagen, *Metafisica,* tr. it. a cura di F.F. Calemi, Cantagalli, Siena 2011, pp. 325-365.

seca. Questa posizione è stata successivamente battezzata come *Turing machine functionalism*, e risulta sintetizzabile, per usare le parole di Dennett, nella seguente maniera:

> per ogni predicato "mentale" "M" esiste un predicato "F" esprimibile in un linguaggio fisicamente neutro, ma progettato per specificare funzioni astratte e relazioni funzionali. I candidati più ovvi per un tale linguaggio sono i sistemi utilizzati per descrivere i computer o i programmi[16].

Nella riflessione sulla filosofia della mente, infatti, il funzionalismo è quella posizione che non richiede la traducibilità sistematica tra predicati mentali e predicati fisicamente neutri. Per il funzionalismo la natura di uno stato mentale non dipende dalla sua struttura interna, ma dalla funzione che svolge e dal ruolo che ricopre all'interno del sistema di cui fa parte:

> Turing identifica i pensieri con stati di un sistema determinati esclusivamente dal loro ruolo nel generare ulteriori stati interni e risposte verbali, a partire da determinati input linguistici. Secondo questa prospettiva [...] diversi sistemi, anche fisicamente molto diversi tra loro, possono avere stati interni [qualitativamente differenti] che svolgono le medesime funzioni[17].

Pertanto sia lo stato interno di una MTU, sia lo stato interno di un computatore umano (costituito da una certa attivazione del suo sistema nervoso), possono svolgere, nonostante l'enorme differenza qualitativa che li contraddistingue, la medesima funzione; e se attribuiamo intelligenza all'output di tale funzione, allora contraddistinto da intelligenza sarà anche l'operatore che l'ha prodotto, chiunque *egli/esso* sia.

Critiche al modello di Turing

Le considerazioni sviluppate da Turing hanno ispirato numerose riflessioni sulla adeguatezza dell'analogia mente-calcolatore. Notoria la posizione di Putnam, tra le principali, che in un'opera del 1961 radicalizza il funzionalismo di Turing:

> Vorrei suggerire che ci sono molte considerazioni che puntano all'idea che una macchina di Turing più elementi casuali sia un modello ragionevole per il cervello umano[18].

---

16 - D. Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology*, Bradford Books, Cambridge 1978, p. XV.

17 - J. Levin, *Functionalism*, in E.N. Zalta - U. Nodelman (a cura di), *The Stanford Encyclopedia of Philosophy*, Summer 2023, Metaphysics Research Lab, Stanford University, 2023.

18 - H. Putnam, *Some issues in the theory of grammar*, in H. Putnam, *Mind, Language and Reality: Philosophical Papers, volume 2*, Cambridge University Press, Cambridge 1975, p. 102.

Tuttavia le critiche che la posizione di Turing ha incontrato sono svariate. Uno dei presupposti cruciali dell'argomento di Turing è che il cervello umano possa essere descritto come un meccanismo a stati-discreti. Come già detto, un meccanismo a stati discreti è un sistema che opera attraverso un numero finito di stati ben definiti, con transizioni tra stati determinate da regole precise e input specifici. Turing afferma che

> Il sistema nervoso non è certamente una macchina a stati-discreti. Un piccolo errore d'informazione sull'intensità di un impulso nervoso che colpisce un neurone può causare un enorme differenza sull'intensità dell'impulso in uscita[19].

Tuttavia Turing appare persuaso che, nonostante la sua apparente complessità biologica e la sua presunta struttura continua, il funzionamento del sistema nervoso possa essere descritto nei termini di un modello finitistico e discreto, in cui segnali e processi sono rappresentabili attraverso stati distinti e regole di transizione[20]. Ma se la fisiologia del sistema nervoso può essere descritta nei termini di una macchina a stati-discreti, e se una MTU è in grado di emulare il comportamento di qualsiasi macchina a stati-discreti, allora una MTU può emulare l'attività del cervello umano. E, come visto, sotto le specifiche condizioni previste dall'esperimento mentale del gioco dell'imitazione, sarebbe perfino possibile attribuire "pensieri" o stati mentali a un dispositivo che, di per sé, non possiede la fisiologia del cervello umano.

Questo punto, come già anticipato, è certamente uno dei più delicati della strategia argomentativa di Turing e indubbiamente anche uno dei più contestati dai suoi critici. Gödel, ad esempio, da un lato non esita a riconoscere i meriti delle implicazioni logico-matematiche dell'opera di Turing, scrivendo quanto segue nella celebre *Gibbs Lecture* del 1951:

> Ci sono vari percorsi diversi per arrivare a questa definizione [di procedimento finito]. Il più soddisfacente, a mio parere, è la riduzione del concetto di procedimento finito a quello di una macchina con un numero finito di componenti, come è stato fatto dal matematico inglese Turing[21]

---

19 - Turing, *Meccanismo computazionale e intelligenza artificiale*, cit., p. 32.

20 - Precisiamo che sebbene Turing sembri assumere che il sistema nervoso operi in maniera continua e non discreta, questa ipotesi non è affatto scontata. Esistono anzi buone ragioni per considerare che, almeno a un certo livello di astrazione, il comportamento neuronale possa essere descritto efficacemente come un sistema a stati discreti. Alcuni modelli di neuroscienze computazionali – così come l'intera architettura delle reti neurali artificiali – si fondano proprio su tale ipotesi.

21 - K. Gödel, *Alcuni teoremi basilari sui fondamenti della matematica e loro implicazioni filosofiche*, in K. Gödel, *Opere. Saggi inediti e conferenze*, tr. it. a cura di G. Lolli - C. Mangione - E. Ballo, Vol. 3, Bollati Boringheri, Torino 2002, p. 268

ed esprimendo, nel passo successivo, forti riserve sulle applicazioni filosofiche di tali innovazioni:

> Per quel che concerne le conseguenze filosofiche dei risultati che esaminiamo, io non credo che siano mai state discusse adeguatamente, o semplicemente [siano state] notate[22].

Per Gödel infatti le teorie riduzioniste della mente trascurano una disgiunzione fondamentale:

> o la mente umana (perfino nell'ambito della matematica pura) supera infinitamente la potenza di ogni macchina finita, oppure esistono proposizioni matematiche assolutamente indecidibili, ossia verità che sfuggono non solo a ogni sistema formale, ma anche a ogni possibile forma di intuizione matematica umana[23].

Tale premessa ha, per Gödel, una conseguenza essa stessa disgiuntiva, ossia o

> il funzionamento della mente umana non può essere ridotto al funzionamento del cervello che secondo ogni apparenza è una macchina finita con un numero finito di componenti, cioè i neuroni e i loro collegamenti[24]

oppure «esistono proposizioni matematiche assolutamente indecidibili» e ciò porterebbe a negare «l'idea che la matematica sia soltanto una nostra creazione»[25]. Benché nel 1951 Gödel non scelga esplicitamente tra le due alternative, mostra comunque una preferenza per la prima, suggerendo che la mente umana possa accedere a verità che nessuna macchina è in grado di generare. Nel 1972 la sua posizione diviene meno sfumata, giungendo a sostenere che

> Turing [...] propone un'argomentazione che si ritiene mostri come i processi mentali non possano andare al di là di quelli meccanici. Tuttavia l'argomentazione non è conclusiva[26].

La preoccupazione di Gödel, infatti, risiede nel fatto che il modello della mente proposto da Turing manifesta un limite fondamentale: l'incapacità di catturare il carattere dinamico e in evoluzione del pensiero umano. Secondo Gödel, invece, la mente non è un sistema statico che opera con un insieme fisso di concetti o strumenti, ma un'entità che si sviluppa costantemente, ampliando sia il numero di termini astratti che è in grado di comprendere sia la precisione con cui questi vengono

---

22 - *Ibid.*, p. 274.

23 - *Ibidem*.

24 - *Ibid.*, p. 275.

25 - *Ibidem*.

26 - K. Gödel. *Alcune osservazioni sui risultati di indecidibilità*, in K. Gödel, *Opere. 1938-1974*. tr. it. a cura di G. Lolli - C. Mangione - E. Ballo, Vol. 2, Bollati Boringheri, Torino 2002, p. 306.

intesi e definiti. Questo processo evolutivo potrebbe essere guidato da metodi sistematici, ma anche in presenza di vincoli finiti in ogni momento, la sua crescita cumulativa può tendere all'infinito:

> anche se a ogni stadio il numero e la precisione dei termini astratti a nostra disposizione può solo essere finito, entrambi (e quindi anche il numero degli stati distinguibili della mente introdotti da Turing) possono nel corso dell'applicazione del processo convergere all'infinito[27].

In sostanza, Gödel sottolinea che la mente ha una capacità potenzialmente illimitata di espandere il proprio orizzonte concettuale, andando oltre i confini rigidi di un modello puramente meccanico come quello immaginato da Turing e codificato dalla MTU.

Analogamente, Nagel e Newman sottolineano, in relazione ai teoremi di incompletezza di Gödel e con un chiaro riferimento implicito a Turing, che

> Il cervello umano [...] sembra incarnare una struttura di regole operative che è decisamente più potente rispetto alla struttura delle macchine artificiali concepite fino ad oggi [...]. La dimostrazione di Gödel [...] implica che le risorse dell'intelletto umano non sono state e non possono essere completamente formalizzate, e che nuovi principi di dimostrazione attendono costantemente di essere inventati e scoperti [...]. Il teorema indica che la struttura e le capacità della mente umana sono molto più complesse e sottili rispetto a qualsiasi macchina non vivente immaginata finora[28].

Non meno importante, in tale quadro, risulta la critica sollevata dal *Chinese Room Argument* di Searle, volta a minare l'idea che un sistema computazionale possa conseguire una vera comprensione del significato delle informazioni che elabora. Searle utilizza il caso ipotetico di una persona che si trova in una stanza e manipola con maestria simboli costituiti da ideogrammi cinesi senza comprenderne il significato, per mostrare che la mera manipolazione di simboli, operata in ottemperanza a manuali di traduzione meramente formali, non implica necessariamente la possibilità di attribuire pensiero o comprensione all'operatore che svolge il processo di manipolazione simbolica[29].

---

27 - *Ibid.*, p. 306.

28 - J. Newman - E. Nagel, *Gödel's Proof*, Routledge, New York 1958, pp. 100-101. Appare peraltro, anche alla luce di quanto indicato fin qui, che la posizione di Nagel e Newman sorvola – o quantomeno lascia sullo sfondo – le considerazioni svolte da Gödel nella *Gibbs lecture* del 1951 sulla natura disgiuntiva delle conseguenze filosofiche dei risultati di incompletezza.

29 - Per un approfondimento rimandiamo a F.F. Calemi, *Significato, pensiero e soggettività: origine e sviluppi della filosofia analitica*, in M. Marianelli *et al.* (a cura di), *Anima, corpo, relazioni. Storia della filosofia da una prospettiva antropologica*, Vol. 3, Città Nuova, Roma 2022, pp. 467-480.

Ricordiamo infine una critica ante-litteram proposta da Wittgenstein sul tema delle macchine pensanti, che anticipa alcune delle questioni sollevate venti anni dopo da Turing. Poco si ricorda, infatti, che la domanda «Una macchina può pensare?» venne già sollevata da Wittgenstein negli appunti che nel 1933-34 dettò ai propri studenti a Cambridge, poi raccolti nel testo pubblicato postumo noto come *Libro blu*. Wittgenstein, nel definire il pensiero come un *operare con segni*[30], anticipa una possibile obiezione: «Quest'obiezione è espressa nella domanda: "Potrebbe una macchina pensare?"»[31]. Come spesso accade nei testi di Wittgenstein scritti dopo il *Tractatus*, a domande si tenta di rispondere sollevando altre domande che mettano in luce l'insensatezza delle prime: «Per ora nominerò soltanto una domanda analoga: "Può una macchina avere il mal di denti?"». Wittgenstein evidenzia che la vera difficoltà del quesito non sta nel fatto che non abbiamo ancora costruito una macchina in grado di "pensare", ma nel fatto che l'usuale grammatica della parola "pensare" viene violata dalla stessa posizione della domanda: quello del *pensare* sarebbe in altri termini un gioco linguistico a cui la macchina non può prendere parte. Nelle parole di Wittgenstein:

> Il grattacapo consiste […] nel fatto che la frase "una macchina pensa" (percepisce, desidera) sembra in qualche modo priva di senso. È come se avessimo chiesto: "Il numero 3 ha un colore?". ("Che colore potrebbe essere, visto che ovviamente non ha nessuno dei colori a noi noti?")[32].

Benché spesso i tentativi di affrancamento da tali critiche si basino sul sottolineare che il motivo dominante dell'opera di Turing sia l'attribuibilità di intelligenza a sistemi computazionali operanti nei contesti specifici previsti dal gioco dell'imitazione, resta il fatto che, sullo sfondo di tale riflessione e del parallelismo funzionale da essa implicato, permanga l'idea della riducibilità computazionale dell'intelligenza umana. Questa concezione, sebbene abbia avuto sostenitori nell'ambito

---

30 - Per motivi di spazio non possiamo approfondire la questione in questa sede e pertanto rinviamo a F.F. Calemi, *Parlare «come gioca uno scacchista». Matematica e semantica degli scacchi in Wittgenstein*, in M. Moschini (a cura di), *Il riconoscimento in gioco*, Pièdimosca, Perugia 2021, pp. 33-53.

31 - L. Wittgenstein, *Libro blu e libro marrone*, Einaudi, Torino 2000, p. 25.

32 - *Ibid.*, pp. 65-66. Si noti come la posizione di Wittgenstein in merito alla domanda «Potrebbe una macchina pensare?» è esattamente la medesima che Turing sposa nel suo articolo 1950: «Penso che la domanda iniziale, "Le macchine possono pensare?" sia tanto priva di significato da non meritare di essere discussa» (A. Turing, *Meccanismo computazionale e intelligenza artificiale*, cit., pp. 19-20). Tuttavia, mentre in Wittgenstein vi è l'intento terapeutico di dissolvere il «grattacapo» filosofico mostrando l'inadeguatezza dell'uso linguistico fuori dal contesto delle forme di vita *umane*, Turing aggira il riferimento a tali forme di vita attraverso una riformulazione operativa del quesito, motivata dall'intento di esplorare le potenzialità del modello rappresentato dalla MTU.

dell'intelligenza artificiale e della filosofia della mente[33], solleva questioni di notevole complessità, sia dal punto di vista epistemologico che ontologico. Infatti, da un lato, la possibilità di modellare il pensiero umano attraverso sistemi computazionali impone di chiarire in che misura la nozione di computazione possa davvero esaurire la ricchezza dei processi cognitivi; dall'altro, la prospettiva riduzionista si scontra con le difficoltà teoriche legate alla coscienza, alla comprensione semantica e alla natura dell'intenzionalità. Come appena visto, la questione rimane altamente problematica e ancora aperta, lasciando spazio a ulteriori indagini e a possibili sviluppi teorici che potrebbero ridefinire i confini tra mente e macchina[34].

Conclusione

Al di là delle aporie da essa sollevate, resta il fatto che la riflessione di Alan Turing rappresenta un punto di svolta nella storia del pensiero scientifico e filosofico, offrendo una nuova prospettiva sulla relazione tra mente, macchina e intelligenza. Il test di Turing rappresenta, come visto, non una mera verifica da sottoporre a un computer ma un invito a riflettere criticamente sul modo in cui attribuiamo significato e valore al comportamento intelligente: la sua eredità si estende ben oltre il campo dell'informatica, ispirando discussioni interdisciplinari che abbracciano filosofia, psicologia, neuroscienze e scienze cognitive.

L'opera di Turing ha quindi aperto la strada a una comprensione più critica e metodologicamente consapevole della natura del pensiero e dell'intelligenza, rappresentando – sia come modello a cui ispirarsi sia come posizione da cui prendere le distanze – un punto di riferimento per ogni interrogazione, oggi più che mai attuale, sul rapporto tra uomo e macchina, natura ed emulazione, pensiero e calcolo[35].

FRANCESCO F. CALEMI
Professore associato di Logica e Filosofia della logica presso il Dipartimento di Filosofia, Scienze sociali, Umane e della Formazione dell'Università degli studi di Perugia
francesco.calemi@unipg.it

33 - Come, ad esempio, W. McCulloch - W. Pitts, *A Logical Calculus of the Ideas Immanent in Nervous Activity*, in «Bulletin of Mathematical Biophysics», 7 (1943), pp. 115-133; H. Putnam, *Some issues in the theory of grammar*, cit.

34 - Per una rassegna puntuale delle principali posizioni che criticano il modello mentale computazionale si veda L. Cimmino, *Test di Turing, intenzionalità e solipsismo*, in A. Turing, *Meccanismo computazionale e intelligenza artificiale*, cit., pp. 51-178.

35 - L'autore desidera ringraziare un referee anonimo per i preziosi commenti forniti su una versione precedente del presente articolo.