

Giulio Amore

Turing filosofo trascendentale: computazione, macchine intelligenti e razionalità

Introduzione

Alan Turing rappresenta una delle più importanti figure scientifiche del secolo scorso. Il numero di articoli in cui Turing viene citato è oggi in costante crescita, soprattutto a partire dal 2020¹. Ciò è sicuramente da mettere in relazione con il recente *AI-boom*, lanciato dal susseguirsi dei vari rilasci di sistemi di intelligenza artificiale (prevalentemente *Large Language Models*), ma è anche una dimostrazione dell'attualità del lavoro del matematico inglese. In questo articolo, intendo sostenere, al pari di Hainscho (Hainscho 2017) e di Hodges (Hodges 1997), come i lavori di Turing sulle macchine intelligenti e sulla computazione possano essere considerati, oltre che matematicamente rilevanti, anche *filosoficamente significativi*. Tuttavia, intendo caratterizzare ulteriormente l'indagine di Turing come *trascendentale*, incentrata da un lato su condizioni di possibilità e limiti trascendentali della computazione e delle macchine intelligenti, dall'altro sull'idea che l'intelligenza o meno di un sistema non dipenda dalla presenza di un soggetto (intelligente) dotato di caratteristiche intrinseche a renderlo tale, ma sulla presenza o meno di attività riconosciute come intelligenti.

Innanzitutto, l'articolo prenderà in esame alcuni importanti lavori di Turing sulle macchine intelligenti e la computabilità (§1), per poi proseguire attraverso il più celebre articolo sull'*Imitation game* (§2). Dopodiché, elencherò le ragioni in base alle quali si può considerare Turing come

¹ Stando alle statistiche di *Google Scholar*, su un totale di 74597 citazioni, ben 25514 (poco più di un terzo) sono posteriori al 2020. *Cfr.* <https://scholar.google.com/citations?user=VWCHlwkAAAAJ&hl=it&oi=sra> (accesso effettuato in data 28/05/2025).

un filosofo trascendentale (§3), esplorando infine la tesi che la computazione stessa possa essere considerata come una forma di razionalità (§4).

1. Turing e le macchine intelligenti: *Entscheidungsproblem* e *Intelligence machinery*

Turing comincia a lavorare all'idea di una intelligenza meccanica sin dal 1941, anno nel quale si dice circolasse un ormai perduto manoscritto sull'argomento (Copeland 2004). Nel 1947 il matematico inglese tenne il suo primo intervento sull'"intelligenza computazionale" e nell'anno seguente scrisse quello che, secondo Jack Copeland, può essere considerato il primo vero "manifesto" dell'intelligenza artificiale (Copeland 2004).

Il saggio, pubblicato postumo, comincia anticipando alcune possibili obiezioni all'ipotesi di una *macchina intelligente*, molte delle quali riconducibili a tre tendenze: un pregiudizio tipicamente umano di non avere rivali in termini di intelligenza; una sorta di *irriverenza prometeica* ("una credenza religiosa che ogni tentativo di costruire tali macchine sia una sorta di Irriverenza Prometeica" (Turing 2004, p. 410)); la convinzione che ogni manifestazione di uno o più comportamenti intelligenti sarebbe solo un mero riflesso del rispettivo creatore (Turing 2004, pp. 410-411). Un'ulteriore obiezione, dice Turing, può essere ricavata dai teoremi di Gödel:

Recentemente, il teorema di Gödel e i risultati ad esso correlati [...] hanno mostrato che, se si tenta di utilizzare delle macchine con scopi come determinare la verità o la falsità di teoremi matematici e non si è disposti a tollerare un errore occasionale, allora qualsiasi macchina data in alcuni casi non sarà in grado di fornire alcuna risposta. Al contrario l'intelligenza umana sembra capace di trovare sempre migliori per risolvere quei problemi, 'trascendendo' i metodi a disposizione delle macchine. (Turing 2004, pp. 410-411)

La presenza o assenza di intelligenza, controbatte Turing, non dipende dalla possibilità di compiere un errore, infatti, essa non risiede nel successo di una azione, ma nell'effettivo *uso* di una facoltà o nell'*esercizio* di determinate funzioni (Turing 2004).

Per testare seriamente l'ipotesi di una macchina intelligente dobbiamo selezionare quelle "*branches of thought*" in cui la macchina possa effettivamente applicarsi e il cui svolgimento sia: 1) subordinato al possesso di una qualche forma di intelligenza; 2) non richiedente necessariamente un corpo completo di organi sensori e motori. Queste *attività* sono: giochi come poker o scacchi, apprendimento linguistico, traduzione, crittografia, matematica (Turing 2004, p. 420).

Tuttavia, è necessario soffermarsi sul tema dell'apprendimento di una macchina (*"Education of machinery"*). Pensare infatti di poter realizzare una macchina che imiti un'intelligenza umana ben formata è un'impresa non solo difficile, ma anche poco realistica. Molto più promettente sembra invece la possibilità di partire da una macchina non-organizzata e di organizzarla via via attraverso un sistema premio-punizione (Turing 2004, pp. 421–422).

Pertanto, dopo alcune proposte preliminari, Turing presenta un tipo particolare di macchina non-organizzata, basata anch'essa su un sistema premio-punizione, ma più raffinato, chiamata *p-type machine*. Questa *p-type machine*, avverte Turing, è da considerare come una *"logical computing machine"* (LCM) senza un vero e proprio nastro che, nella sua versione "universal", rappresenta quella che viene ricordata come *Macchina di Turing* (TM) (Turing 1936). Turing, riassumendo, descrive una LCM così:

[Ha] una capacità di memoria infinita, realizzata sotto forma di un nastro infinito suddiviso in caselle, ciascuna delle quali poteva contenere un simbolo. In ogni momento vi è un solo simbolo all'interno della macchina, detto simbolo scansionato. La macchina può modificare il simbolo scansionato, e il suo comportamento è in parte determinato da esso; tuttavia, i simboli presenti altrove sul nastro non influenzano il comportamento della macchina. Il nastro, però, può essere fatto scorrere avanti e indietro attraverso la macchina: questa è una delle operazioni elementari che essa è in grado di compiere. Di conseguenza, qualsiasi simbolo presente sul nastro potrà, in linea di principio, entrare in gioco. (Turing 2004, p. 413)

È alla luce dell'articolo del 1948 che diventa esplicita l'importanza fondamentale, anche in campo cognitivo, del concetto di TM proposto nel 1936, testimoniando la stretta vicinanza ed estrema continuità dei due articoli (Shanker 1987). Sintetizzando la conclusione dell'articolo del 1936, possiamo asserire che: un numero è computabile quando può essere scritto e trasformato attraverso un set di regole pre-determinato da una macchina (Turing 1936). Cos'è concretamente tuttavia la *computazione*? Se un problema è computabile, significa che è risolvibile da una TM, ovvero, è risolvibile da una macchina ideale, con memoria infinita, che, attraverso permutazioni guidate da una o più regole di esecuzione, manipola i simboli a sua disposizione per raggiungere una soluzione. La computazione, pertanto, sarà il processo mediante il quale un problema, rappresentabile attraverso simboli, viene risolto in base a determinate regole di esecuzione formalizzabili. Turing, introducendo il concetto di *computazione* per risolvere l'*Entscheidungsproblem*, apre le porte ad un'ipotesi che tutt'oggi mostra ancora la sua proficua problematicità: se un'operazione è computabile, allora è risolvibile da una macchina.

L'articolo del 1948 mostra come TM non sia solamente un artefatto concettuale riguardante la sola matematica, atto a risolvere il problema della computabilità, ma un artefatto concettuale che riguarda da vicino filosofia e tecnologia, che mette in crisi un concetto di intelligenza antropocentrico e che sulla cui crisi permette di costruire l'ipotesi di una macchina intelligente. L'idea di intelligenza intesa a partire da una facoltà riservata esclusivamente all'essere umano viene definitivamente affrontata e messa in discussione nel celebre articolo del 1950: *Computing Machinery and Intelligence*.

2. Il ruolo della *performance* nell'*Imitation Game*

L'articolo, considerato il primo ad affrontare la possibilità di meccanicizzare l'intelligenza umana (Nilsson 2010), comincia sin da subito col mettere in discussione la domanda a cui parrebbe fondamentale dare una risposta, essendo all'interno di un dibattito sulle macchine pensanti. La domanda "*Can machines think?*" è infatti considerata come ambigua. Come rispondere se non abbiamo infatti chiaro né cosa significhi il termine *macchina* né *pensare*? (Turing 1950, p. 433).

Turing, pertanto, sposta il focus della domanda o, meglio, sposta il terreno di risposta. Non si tratterà di elaborare una teoria del pensiero che possa per così dire accomodare una o più definizioni di macchina, di computazione e di intelligenza macchinica, bensì, di elaborare un gioco, l'*Imitation game*, che passerà alla storia anche come *Test di Turing*, che ci permetta di riflettere sull'interrogativo iniziale sulla base di un comportamento, di una *performance*. Questo è un elemento fondamentale e discriminante: ancora una volta, l'approccio di Turing è basato su una *performance* intellettuale, non sulla struttura del cervello umano o di quella "neurale" della macchina.

Nelle pagine conclusive dell'articolo, egli sembra concludere che per vedere effettivamente dei risultati positivi, dunque, diremo noi, degli epifenomenici comportamenti cosiddetti intelligenti, bisognerà attendere quantomeno la fine del secolo (Turing 1950, p. 442). Tuttavia, l'obiettivo di Turing non è programmare nel 1950 una macchina intelligente, ma tracciare le condizioni di possibilità di una macchina che manifesti, performativamente parlando, un comportamento intelligente agli occhi di un osservatore esterno e analogo ad uno dello stesso tipo performato da un essere umano e per fare ciò propone il noto *Imitation Game* (Turing 1950, p. 434)².

² Per un resoconto dettagliato dell'*Imitation Game*, vedi anche (Grigenti 2021, pp. 185-189)

Ora, dice Turing, se sostituissimo uno dei partecipanti e mettessimo una macchina al suo posto, cosa accadrebbe? (Turing 1950, p. 434).

La macchina da introdurre all'interno del gioco, tuttavia, non sarà una macchina qualsiasi, bensì una macchina che Turing chiama *digital computer*. Non esisteva un *digital computer* all'epoca di Turing, ma non c'è bisogno che fosse esistito realmente, che il test, diremmo, fosse concretamente effettuato; l'importante è che ci fossero (e che ci siano) le condizioni di possibilità per poter immaginare una macchina capace di superarlo, capace di ingannare l'interrogatore.

[Noi] non ci stiamo chiedendo se tutti i computer digitali si comporterebbero bene nel gioco o se i computer attualmente a disposizione lo farebbero, ma se esistano computer concepibili che lo farebbero. (Turing 1950, p. 436)

La via tracciata, incentrata non tanto sulla vera e propria progettazione di una macchina intelligente, quanto più sull'idea di una macchina ipotetica con la possibilità di svolgere compiti intellettualmente caratterizzati in modo analogo all'essere umano, sfocia nel paragrafo 5, intitolato *Universality of Digital Computers*.

Innanzitutto, il *digital computer* può esser considerato una '*discrete state machine*' (oggi diremmo a stati finiti, o *Finite-state machine*). Nello specifico, questo implica che la macchina passi da uno stato definito all'altro, in base a delle regole. Conoscendo le regole di azione, a partire da un qualsiasi stato discreto si può prevedere il successivo.

Dunque, una macchina universale ideale a stati discreti, dotata delle corrette tavole di istruzioni, può in teoria replicare il comportamento di una qualsiasi specifica macchina a stati discreti. Perciò, se il *digital computer* può effettivamente calcolare, sulla base di uno stato *x*, un qualsiasi stato *y* successivo, determinato in base allo specifico set di istruzioni a disposizione, va da sé che, in virtù del fatto di poter operare con qualsiasi set di istruzioni, e di conseguenza imitare ogni macchina a stati discreti, esso possa essere una *macchina universale* (Turing 1950, p. 441).

La domanda iniziale (*Can machines think?*) è dunque definitivamente riformulata nel seguente modo:

Possiamo ora tornare a considerare la questione sollevata alla fine del §3. Era stato suggerito, in via ipotetica, di sostituire la domanda "Le macchine possono pensare?" con "Esistono computer digitali concepibili che si comporterebbero bene nel gioco dell'imitazione?". Se lo si desidera, si può formulare questa domanda in modo apparentemente più generale e chiedere: "Esistono macchine a stati discreti che si comporterebbero bene?". Tuttavia, alla luce della proprietà di universalità, si comprende che entrambe queste formulazioni sono equivalenti alla seguente: "Fissiamo la nostra attenzione su un particolare computer digitale *C*. È vero che, modificando opportuna-

mente questo computer per dotarlo di una memoria adeguata, aumentando sufficientemente la sua velocità di esecuzione e fornendogli un programma appropriato, C possa essere reso capace di svolgere in modo soddisfacente il ruolo di A nel gioco dell'imitazione, assumendo che il ruolo di B sia ricoperto da un essere umano?" (Turing 1950, p. 442)

L'idea di fondo è che una conversazione effettuata attraverso un sistema tale da impedire di avere contatto diretto tra gli interlocutori, il che rivelerebbe evidentemente l'identità degli stessi, sia effettivamente "performabile" da una macchina, nello specifico un *digital computer*, in maniera sufficientemente adeguata da trarre in inganno l'interrogatore, sembrando ai suoi occhi sufficientemente umano.

Nel settimo e ultimo capitolo, intitolato *Learning Machines*, Turing riprende alcune riflessioni che abbiamo già potuto analizzare nella disamina dell'articolo inedito del 1948. In questa sede Turing propone un'alternativa alla programmazione, fin troppo complessa, di una macchina che simuli una mente adulta: produrne una che simuli la mente di un bambino e che apprenda e si strutturi a partire da una programmazione iniziale meno impegnativa, orientata all'apprendimento.

Come nello scritto del 1948, anche qui il metodo fondamentale di insegnamento viene riconosciuto nel sistema premio-punizione, accompagnato tuttavia da uno sfondo di inferenze logiche, fatti e congetture riconosciute come vere.

Tuttavia, lo stesso Turing intravede la direzione per una svolta che permetta di andare oltre questi "imperativi" logici da inscrivere all'interno del codice di un programma: inserire un elemento di casualità nella macchina ("*a random element in a learning machine*") per quei problemi che presentano un ventaglio di soluzioni accettabili, al fine di replicare un comportamento intelligente che non consiste in un comportamento completamente casuale, ma nemmeno perfettamente disciplinato (Turing 1950, p. 459).

Sganciarsi leggermente da un comportamento perfettamente disciplinato, introdurre un elemento *casuale*, riuscire a trovare una soluzione sufficientemente adeguata (e non quella assolutamente migliore, nel caso di più soluzioni valide), introdurre l'idea di un processo che sia rappresentabile da una *percentuale* di certezza, dunque non assoluta, sono tutti passaggi molto chiari nella direzione di un tipo di programmazione che dovrà attendere più di trent'anni prima di essere effettivamente realizzabile, ma che dimostrerà di essere probabilmente l'approccio migliore (Gershman et al. 2015). Inoltre, sia l'idea di un sistema "bambino" che si forma da sé, che l'idea di introdurre un elemento di casualità all'interno della macchina diventano ancora più significative se considerate alla luce del successivo lavoro di Turing sulla *morfogenesi* (Turing 1952) e sulla rilevanza in esso dei concetti di *forma* e *formazione* (Grigenti 2022).

Concludendo, il concetto di *macchina intelligente* presente nei lavori di Turing indica una macchina in cui l'atto-computativo rappresenta l'atto cognitivo, in cui la programmazione diventa lo strumento chiave di successo, in cui lo scopo è performativo: fare sì che la macchina funzioni in maniera analoga alla mente umana³, basandosi sull'idea, tipicamente wittgensteiniana⁴, che la capacità cognitiva si manifesti in attività e non in forme certe e stabili (Tamborini 2025).

3. Turing come filosofo trascendentale

Kant definisce *trascendentale* quel tipo di conoscenza a priori che si occupa del modo di conoscere gli oggetti e non tanto degli oggetti in quanto tali (Kant 1968, B25). Turing si interroga direttamente sulle condizioni di possibilità della computazione stessa, definendone in linea di principio le condizioni affinché possa verificarsi un "atto computativo" in quanto tale (i.e. se il problema è risolvibile da una macchina universale). Come fa notare Hodges (Hodges 1997), la soluzione al problema della "dimostrabilità" di Hilbert presentata nell'articolo del 1936 è profondamente originale, pur giungendo alle stesse conclusioni (negative) di Church (Church 1936).

Turing, infatti, abbandona la tradizionale dimostrazione matematica e presenta invece quello che a tutti gli effetti è un esperimento mentale, tipico strumento filosofico (Casati 2011), introducendo una nuova idea,

³ La rilevanza per quanto riguarda questo aspetto di Turing viene sottolineata anche da McCorduck (McCorduck 2004, 143, in nota).

⁴ Sono note le critiche mosse da Wittgenstein a Turing come filosofo o, meglio, di come Wittgenstein non accettasse l'argomento-ponte tra matematica e filosofia della mente rappresentato dalla *TM* presente nell'articolo del 1936 (Shanker 1987). Le critiche di Wittgenstein erano dirette in generale all'idea stessa di macchine intelligenti, considerate incapaci di produrre una vera e propria *prassi* intelligente in virtù della loro stretta rigidità e della loro aderenza assoluta alle regole di programmazione. Wittgenstein, infatti, pensava che le *TM* fossero di fatto un'estrema formalizzazione di un "uomo-che-calcola" (Wittgenstein 1980). Sebbene molti critici considerino questi elementi per tracciare un confine netto tra le due figure, financo un conflitto, a volte traslasciano la necessità, ravvisata invece da Shanker (Shanker 1987), di contestualizzare cronologicamente queste critiche, avvenute in gran parte quando ancora l'idea di *macchine intelligenti* di Turing non era pienamente sviluppata. Inoltre, proprio l'attenzione di Turing alla *performance*, all'attività, alla possibilità di un apprendimento guidato da un sapere che si forma con il rispetto di "regole" ma solo all'interno della "prassi", non può non far pensare alle pagine delle *Ricerche* riguardanti l'apprendimento tramite rispetto e contravvenzione alle determinate regole e comandi validi in un contesto, in un *gioco* (Wittgenstein 1953). Per una disamina più approfondita delle possibili interpretazioni del rapporto tra il pensiero di Wittgenstein e Turing cfr. (Proudfoot 2024).

un nuovo concetto, quello di computabilità. Inoltre, all'interno del suo articolo, Turing mostra un interesse particolare per i limiti della computazione, il che caratterizza ancora di più la proposta in termini trascendentali classicamente intesi.

La ricerca di una soluzione all'*Entscheidungsproblem* (dimostrare se ogni problema formalizzato sia deducibile attraverso un sistema formale e dunque universalmente valido) diventa la ricerca di una soluzione al problema trascendentale della computazione (poter stabilire a priori se un enunciato sia o meno deducibile). La risposta, però, è altrettanto trascendentale, poiché individua le condizioni di possibilità ed i limiti della computazione (per esempio quello denominato successivamente come *Halting Problem* (Rogers (Jr.) 1957)), rispondendo così di fatto anche alla domanda originaria: siccome esistono problemi non computabili, esistono problemi non risolvibili attraverso un sistema formale di calcolo.

Ma come stabilire quali operazioni siano computabili? La risposta viene data attraverso la proposta di un "*criterio attivo di decidibilità*": sono computabili quelle funzioni che possono essere computate da una TM. L'elemento radicalmente trascendentale, per certi versi diverso rispetto a quello kantiano già menzionato poco sopra, risiede proprio in ciò: la TM presa così in sé non è una condizione a priori della computabilità (c'è una TM $\rightarrow$ c'è computabilità $\rightarrow$ il problema è risolvibile formalmente), ma è condizione attiva a priori della computabilità (c'è una TM $\rightarrow$ l'attività di quella TM, un processo computativo, può risolvere il problema $\rightarrow$ il problema è risolvibile formalmente). Lo scarto concettuale rappresentato dal costrutto ideale in cui consiste la TM mostra la profonda radicalità dell'argomento di Turing, il quale risolve l'*Entscheidungsproblem* attraverso la soluzione della questione, trascendentale, sui criteri a priori per stabilire, altrettanto a priori, le condizioni di deducibilità formale di un problema. Ma è qui, ribadisco, il passaggio fondamentale: la TM non è un criterio formale a priori, la cui esistenza o semplice presenza basta a soddisfare le condizioni di possibilità della computazione, ma è un criterio attivo a priori la cui attività operativa – se possibile – funge da condizione a priori della possibilità della computazione, e perciò della deducibilità formale del problema in questione. Il fatto che l'attività della TM sia trascendentalmente limitata, inoltre, risolve il problema posto da Hilbert in maniera critico-trascendentale, registrando appunto i limiti dell'attività della TM⁵, quindi della computabilità, concludendo che non ogni problema formalmente esprimibile sia altrettanto formalmente risolvibile.

⁵ E di fatto della TM stessa essendo un costrutto ideale che si dà nel suo essere intesa a partire dalla possibilità di una attività: la computazione.

Il tema dell'attività diverrà centrale in tutti i successivi lavori, culminando nell'articolo del 1950, che rappresenta pertanto il punto di arrivo di questa proposta trascendentale più profonda, radicata nell'idea di un'attività come orizzonte di possibilità della computazione, dunque dell'intelligenza macchinica, non nella presenza o meno di una determinata entità o soggetto già formato⁶. Il focus, infatti, non è sulla macchina universale in quanto tale, ma sull'attività della macchina universale. A fondamento sta un'attività, non la presenza di un principio statico a priori. La vera domanda implicita nell'articolo del 1950 è se effettivamente una “certa capacità di manipolare simboli in un contesto nel quale lo scambio di significati avviene in modo puramente funzionale” (Grigenti 2021, p. 187), i.e. nel caso della macchina uno o più atti computativi, possa fondare l'agire intelligente. Tuttavia, che il *Test* proposto nell'articolo riguardi solo l'agire intelligente di una macchina, non è per nulla scontato, o quanto meno, rimane non precisato e rappresenta ancora oggi una delle maggiori difficoltà interpretative (Grigenti 2021).

Ad ogni modo, Turing opera uno spostamento della domanda stessa: non si chiede più “che cosa” sia intelligente, ma “come” qualcuno o qualcosa possa esserlo.

Nel prossimo paragrafo non analizzerò l'ampio concetto di intelligenza, ma considererò una componente prominente di essa: la razionalità (Burgoyne et al. 2023), cercando poi di dimostrare che la computazione può essere considerata come una forma di razionalità. Se così fosse, infatti, Turing avrebbe posto le basi non solo ad un'interrogazione trascendentale sulla computabilità, ma ad una interrogazione trascendentale *tout court* della razionalità (e di conseguenza anche dell'intelligenza *simpliciter* o quantomeno di una sua componente).

4. Computazione e razionalità

Si potrebbe obiettare che per rispondere alla domanda di cui sopra, sia necessario fornire una definizione precisa di razionalità. Turing stesso evidenzia tuttavia che richieste simili (analoghe al chiedersi cosa sia “pensiero”) sono questioni non solo scomode da affrontare, ma anche possibilmente deleterie al fine di raggiungere delle conclusioni proficue (Turing 1950).

Se prendiamo tuttavia un'attività qualsiasi e proviamo a considerarla sotto due aspetti differenti, affinché dall'uno emerga come un processo

⁶ Un trascendentale che definirei più fichtiano che kantiano, in analogia con l'idea fichtiana di un soggetto che non costituisce il principio primo dell'attività “visiva”, ma che al contrario è fondato immanentemente dall'attività “visiva” stessa all'interno tuttavia di un processo non rappresentabile causalmente (Rametta 2014).

non razionale e dall'altro come processo razionale, potremmo trovare alcuni criteri per individuare le caratteristiche dell'esercizio della razionalità. Consideriamo pertanto il seguente esperimento mentale.

Abbiamo due agenti (A e B) che si ritrovano a dover affrontare lo stesso problema: attraversare un ponte per recapitare un messaggio urgente. Tuttavia, il ponte in questione è stato danneggiato da una tempesta ed è ora pericolante.

A agisce nella seguente maniera: esamina lo stato del ponte, dividendo la struttura in differenti aree per valutare meglio i danni, e nota che alcune travi sono particolarmente danneggiate; ricorda di aver studiato a scuola alcuni principi di costruzione e tenuta dei ponti e li confronta con i danni riscontrati; inferisce che il ponte, se attraversato normalmente, potrebbe non tenere. Considera anche alcune alternative, come guadare il fiume oppure allungare il cammino per attraversarlo dove le sponde sono più vicine. Purtroppo, si rende conto che queste alternative prenderebbero troppo tempo. Decide così di attraversare comunque il ponte, ma appoggiando il peso, distribuendolo mettendosi a pancia in giù se possibile, solo su quelle travi che dovrebbero tenere, in base al suo ragionamento. Nonostante il timore di cadere e annegare, A riesce ad attraversare il ponte senza che crolli.

B si ritrova di fronte alla stessa situazione, tuttavia, agisce in maniera completamente diversa. B pensa che tutto sommato è stato fortunato ultimamente (anche se ha perso dieci franchi il giorno prima scegliendo croce al lancio di una moneta); fa qualche scongiuro a voce alta, poi corre dall'altra parte e una volta arrivato sull'altra sponda esclama: "l'avevo detto che sono fortunato ultimamente!"

B, inconsapevolmente, è riuscito a passare indenne il ponte proprio perché nell'attraversarlo ha appoggiato il suo peso solo su quelle travi portanti, quelle stesse che A aveva individuato grazie alle sue deduzioni.

La stessa azione, attraversare il ponte, è stata eseguita con successo da entrambi gli agenti, tuttavia, l'esperimento mentale mostra delle differenze abissali tra i due tentativi. Difficilmente diremmo razionale la strategia impiegata da B, quanto piuttosto basata su una sua sensazione, giustificabile tra l'altro solo da una superflua ripetizione incapace di spiegare la scelta (perché hai passato il ponte? "Perché sono fortunato." Perché sei fortunato? "Perché sono fortunato.").

A, invece, potrebbe riapplicare la stessa strategia su tutti i ponti che lo permettessero, e sarebbe inoltre capace di spiegare la ragione di ogni singolo passaggio, finalizzato alla prosecuzione in avanti lungo il ponte.

Dunque, la razionalità potrebbe essere intesa come quella capacità che, se usata, permette ad un'entità di dare una risposta ad uno stimolo generalmente complesso, all'interno di un determinato contesto C, applicando un metodo o una combinazione di metodi. La risposta, tuttavia,

non può essere genericamente caratterizzata, altrimenti qualsiasi sistema in grado di rispondere – in qualunque modo – a qualunque forma di eccitazione dovrebbe essere considerato razionale, il che è evidente paradossale (un lucchetto si apre quando viene inserita la combinazione giusta, ma non considereremmo l'azione di aprirsi come razionale). Pertanto, sulla scorta di quanto analizzato precedentemente, possiamo precisare e aggiungere che la risposta deve avere determinate caratteristiche.

Essa deve essere:

Giustificabile: effettuata seguendo un metodo, una o più regole il cui rispetto permetta di rendere ragione della risposta effettuata (dove una forma di coerenza interna). In generale, non casuale nel suo processo.

Produttiva: applicando determinate regole in risposta ad un dato (lo stimolo) semplice o complesso, produce una risposta. È operativa poiché appunto opera, a partire da un materiale di partenza, determinati passaggi, applica regole, e soprattutto produce una risposta.

Contestualizzata: se infatti il processo di risposta non tenesse conto di un criterio di coerenza esterna, se potesse darsi in forma completamente de-contestualizzata, non sarebbe di fatto coerente con lo stimolo e di conseguenza con quel contesto in cui è chiamato a dare risposta (se alla domanda “Quanto fa $2+2$? Si rispondesse: “gelato”, probabilmente non potremmo parlare di razionalità. La risposta, infatti, sarebbe realmente una risposta che attraverso determinate regole è organizzata e predisposta per rispondere a quello stimolo?)

Teleologica: proprio perché è orientata a produrre una risposta coerente, giustificabile, contestualizzata ad un determinato stimolo sulla quale opera, non può essere fine a sé stessa, ma è sempre, e ciò consegue evidentemente dalle caratteristiche sopra citate, orientata ad un fine. Inoltre, il criterio contraddice alcune critiche relative al tema delle “risposte sbagliate”. Immaginiamo che un *Sistema-a* produca una risposta r ad un quesito x . Ipotizziamo che $r(x)$ sia una risposta sbagliata al quesito x in questione. Quella risposta apparirà *a posteriori* come non-giustificabile, mettiamo caso per una contraddizione interna. Ora, se la razionalità del *Sistema-a* non fosse teleologicamente orientata, come potremmo noi affermare che, quando si danno risposte sbagliate, si sia comunque fatto un tentativo? Se non fosse guidata da un fine, ovvero produrre una risposta con le determinate caratteristiche sopra elencate, allora la razionalità verrebbe utilizzata soltanto quando si danno risposte corrette. Ma l'effettivo successo dell'applicazione della facoltà non può essere stabilito *a posteriori* rispetto all'effettivo successo. È il tentativo di dare una risposta giustificabile, coerente, contestualizzata che determina la presenza o meno di una forma di razionalità in atto. In altre parole, è solo se definiamo la razionalità come teleologicamente orientata – orientata alla produzione di una risposta giustificabile, coe-

rente, contestualizzata ad un determinato stimolo – che realmente comprendiamo la facoltà nel suo darsi, nella sua attività, e non riducendola alla correttezza o meno dei suoi risultati.

Considerato ciò, proviamo a spostarci e a valutare se effettivamente la computazione sia o meno una forma di razionalità. Esaminerò ognuna delle precedenti caratteristiche, ipotizzando per assurdo che esse non siano parti del processo computativo.

Produttività: stando alla definizione di Turing, un processo computativo è un processo che può risolvere un problema computabile, ovvero risolvibile mediante una o più permutazioni su un nastro ipotetico infinito attraverso l'applicazione di un determinato set di regole. Per non essere produttiva, la computazione non dovrebbe risolvere problemi, poiché se li risolvesse, produrrebbe una risposta. Ma la computazione è un processo atto a risolvere problemi. Se non dovesse fornire risposte, non avrebbe ragione di esistere. Dunque, se la computazione esiste, essa produce risposte, ed è perciò produttiva. Ci sono problemi a cui però la computazione non può rispondere. Se un problema non ha risposta mediante computazione, allora non è computabile, ma questo non implica che la computazione non sia produttiva, semplicemente pone dei limiti trascendentali alla computazione.

Giustificabilità (e non casualità): per definizione la computazione è quel processo atto alla soluzione di un problema attraverso una o più permutazioni effettuate secondo un set di regole. Se fosse casuale, allora non starebbe seguendo un set di regole, il che sarebbe in contraddizione con la definizione stessa di computazione.

Contestualizzazione: la computazione è per definizione un processo atto alla soluzione di un problema, dunque, è orientata alla soluzione di *quel* problema, il quale rappresenta di fatto il contesto operativo del processo computazionale.

Teleologicità: la computazione è per definizione un processo atto alla soluzione di un problema, tant'è che l'insieme dei problemi può essere suddiviso in problemi computabili e non computabili. Se non fosse orientata ad un fine, non sarebbe un processo atto alla soluzione di un problema, il che va in contraddizione con la sua stessa definizione.

L'idea che processi computativi siano coinvolti all'interno dell'agire razionale umano è al centro di alcune proposte teoretiche di *computation rationality* (Lewis et al. 2014; Gershman et al. 2015). Il tentativo percorso in questo paragrafo, sulla scia del pensiero di Turing, intende andare oltre l'idea di processi razionali computativi circoscritti all'ambito dell'uso della razionalità da parte degli esseri umani e intende proporre, sulla base del precedente esperimento mentale e della successiva dimostrazione, che la computazione – in quanto tale – sia una *forma di razionalità*.

Conclusioni

Nel presente articolo ho cercato di evidenziare, attraverso l'analisi dei principali scritti di Turing su computabilità e macchine intelligenti, come l'indagine del matematico inglese possa essere considerata, filosoficamente, come trascendentale.

Non solo essa affronta il tema delle condizioni di possibilità della computazione, e dei suoi limiti, ma affronta l'argomento suddetto e quello delle macchine intelligenti considerando come l'aspetto operativo, attivo, performativo fondi da un lato le condizioni di possibilità della computazione e dall'altro dell'intelligenza macchinica. Sulla scorta dello stretto legame tra computazione e macchine intelligenti ho cercato di approfondire un possibile percorso di indagine riguardante l'ipotesi che la computazione rappresenti una forma di razionalità, concludendo che essa, proprio perché considerata come facoltà attiva, lo sia.

Infine, questa ipotesi vuole aprire un ulteriore dibattito, riassumibile in due domande finali: se la razionalità, in una delle sue forme, può essere computabile, può parimenti esserlo l'intelligenza? E se così fosse, potremmo dire che Turing avrebbe aperto una nuova stagione di *rivoluzione copernicana* della ragion pura?

Bibliografia

- Burgoyne, A.P., Mashburn, C.A., Tsukahara, J.S., Hambrick, D.Z., Engle, R.W. (2023). *Understanding the relationship between rationality and intelligence: a latent-variable approach*, in «Thinking & Reasoning», vol. 29, n. 1, pp. 1–42.
- Casati, R. (2011). *Prima lezione di filosofia*, Laterza, Roma-Bari.
- Church, A. (1936). *A note on the Entscheidungsproblem*, in «The Journal of Symbolic Logic», vol. 1, n. 1, pp. 40–41.
- Copeland, B.J. (2004). *The Essential Turing. Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence*, Oxford University Press, New York.
- Gershman, S.J., Horvitz, E.J., Joshua, B.T. (2015). *Computational rationality: A converging paradigm for intelligence in brains, minds, and machines*, in «Science», vol. 349, n. 6245, pp. 273–278.
- Grigenti, F. (2021). *Le macchine e il pensiero*, Orthotes, Napoli-Salerno.
- Grigenti, F. (2022). *Alan Turing. Morfogenesi*, in «estetica. studi e ricerche», n. 1/2022.
- Hainscho, T. (2017). *Alan Turing's models of cognition*, https://thweb.at/includes/files/hainscho_turing_koeln_09-2017.pdf
- Hodges, A. (1997). *Turing*, Phoenix, London.
- Kant, I. (1968). *Kritik der reinen Vernunft*, W. de Gruyter, Berlin.
- Lewis, R.L., Howes, A., Singh, S. (2014). *Computational Rationality: Linking Mechanism and Behavior Through Bounded Utility Maximization*, in «Topics in Cognitive Science», vol. 6, n. 2, pp. 279–311.

- McCorduck, P. (2004). *Machines who think: a personal inquiry into the history and prospects of artificial intelligence*, A.K. Peters, Natick (MA).
- McCulloch, W.S., Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*, in «The bulletin of mathematical biophysics», vol. 5, n. 4, pp. 115–133.
- Nilsson, N.J. (2010). *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge University Press.
- Proudfoot, D. (2024). *Wittgenstein and Turing on AI: Myth versus Reality*, in Ball, B., Helliwell, A.C., Rossi, A. (a cura di), *Wittgenstein and Artificial Intelligence*, Volume I: Mind and Language, Anthem Press.
- Rametta, G. (2014). *Fichte: La Logica Trascendentale Come Logica Del Senso*, in «Rivista di Storia della Filosofia (1984-)», vol. 69, n. 4, pp. 737762.
- Rogers (Jr.), H. (1957). *Theory of Recursive Functions and Effective Computability*, Massachusetts Institute of Technology.
- Shanker, S.G. (1987). *Wittgenstein versus Turing on the nature of Church's thesis*, in «Notre Dame Journal of Formal Logic», vol. 28, n. 4.
- Tamborini, M. (2025). *The ethics of bioinspired animal-robot interaction: A relational meta-ethical approach*, in «Journal of Responsible Technology», vol. 22, pp. 100116.
- Turing, A. M. (1936). *On Computable Numbers, with an Application to the Entscheidungsproblem*, in «Proceedings of the London Mathematical Society», vol. s2-42, n. 1, pp. 230–265.
- Turing, A. M. (1950). *Computing Machinery and Intelligence*, in «Mind», vol. LIX, n. 236, pp. 433–460.
- Turing, A. M. (1952). *The Chemical Basis of Morphogenesis*, in «Philosophical Transactions of the Royal Society of London», vol. 237, n. 641, pp. 37–72.7
- Turing, A. M. (2004). *Intelligent Machinery*, in Copeland, B.J. (a cura di), *The Essential Turing. Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence*, Oxford University Press, New York, pp. 395–432.
- Wittgenstein, L. (1953). *Philosophische Untersuchungen*, Basil Blackwell, Oxford.
- Wittgenstein, L. (1980). *Remarks on the Philosophy of Psychology*, Basil Blackwell, Oxford.

Turing filosofo trascendentale: computazione, macchine intelligenti e razionalità

In questo articolo si intende mostrare come le opere di Turing, in particolare quelle dedicate alle macchine intelligenti e alla computazione, siano non solo filosoficamente rilevanti, ma anche interpretabili in chiave trascendentale.

L'analisi prende in esame alcuni principali scritti di Turing, evidenziando la costante attenzione ai temi dell'attività in quanto tale e alla possibilità di comprendere l'intelligenza a partire dal punto di vista della performance. A partire da questa analisi, il saggio affronta due concetti di trascendentale presenti nell'opera di Turing: uno radicato nella *Critica della ragion pura* di Kant, e l'altro fondato sulla presupposizione del ruolo trascendentale dell'attività nella formazione dell'intelligenza o la razionalità. Sviluppando ulteriormente questa analisi, l'articolo esplora l'ipotesi secondo cui la computazione costituirebbe esso stessa una forma di razionalità. Attraverso un esperimento mentale, si individuano quattro caratteristiche dell'attività razionale: giustificabilità, produttività, contestualizzazione e orientamento teleologico. Si dimostra come i processi computazionali possiedano necessariamente tutte e quattro queste caratteristiche, suggerendo che la computazione non si limiti a partecipare ai processi razionali, ma costituisca essa stessa una forma di razionalità. Questa interpretazione colloca Turing interprete di una nuova forma di indagine trascendentale sulla razionalità e sull'intelligenza.

PAROLE CHIAVE: Turing, filosofia trascendentale, computazione, intelligenza artificiale, razionalità.

Turing as transcendental philosopher: computation, intelligent machinery and rationality

This article argues that Alan Turing's works, especially those on intelligent machines and computation should be understood as philosophically relevant and should be interpreted as transcendental.

The analysis examines some of Turing's major works, highlighting the consistent focus on activity and on the possibility of understanding intelligence from the perspective of performance. From this analysis, the paper addresses two concepts of transcendental involved in Turing's work: one grounded in Kant's *Critique of Pure Reason* and the other based on the presupposition of the transcendental role of activity in the formation of intelligence or rationality. Building on this analysis, I delve further into the hypothesis that computation itself constitutes a form of rationality.

Through a thought experiment, I establish four characteristics of rational activity: justifiability, productivity, contextualization, and teleological orientation. I demonstrate that computational processes necessarily possess all four characteristics, suggesting that computation is not merely involved in rational processes but constitutes a form of rationality itself.

This interpretation positions Turing as an interpreter of a new transcendental inquiry into rationality and intelligence.

KEYWORDS: Turing, transcendental philosophy, computation, artificial intelligence, rationality.