

AUTOCOSCIENZA

Preferiremmo tralasciare lo studio dell'autocoscienza e sostituirlo con una serie di questioni sulla possibilità dell'intelligenza artificiale. Questo perché il problema, oltre ad avere radici filosofiche profonde e complesse, trova agganci con tutto quanto trattato finora e rappresenta in certo modo il cuore della ricerca sull'IA, sebbene non ci sia alcun progetto che abbia come suo scopo precipuo la produzione di un computer autocosciente.

Vedremo che le questioni poste, pur partendo da ciò che abbiamo esposto in queste pagine, toccheranno spesso più o meno evidentemente il tema del sapere di sapere e del sapere di essere, oltre alla capacità di trascendersi e autocomprendersi, di cui abbiamo accennato in precedenza.

Struttureremo questa parte proponendo 10 tesi a favore dell'IA e 10 risposte che le mettono in dubbio.

1. L'IA è possibile nella sua ipotesi debole in quanto quella parte dell'attività mentale che viene ritenuta non algoritmica è facilmente implementabile tramite procedure euristiche (cioè tramite l'introduzione di fattori pseudo-casuali nei calcoli) e moduli di logica fuzzy, seppur con alti costi tecnici.

Assunta questa ipotesi della matematizzabilità anche di molteplici varianti in deterministiche, il problema è quello di vedere il loro possibile numero. Se esso non è infinito e quindi indefinibile, di nuovo, l'IA ricadrebbe sotto la prospettiva di un calcolatore comunque programmato tecnologicamente.

Infatti l'ipotesi debole afferma che la macchina può emulare, ma non simulare, e quindi tale problema, della equivalenza di un sistema finito con una macchina, intacca l'ipotesi forte, ma non quella qui sostenuta che rimane valida, seppur di non notevole importanza.

2. La pretesa di uniformità delle procedure della macchina ai processi mentali è inconsistente (e quindi non fornisce un valido metro di giudizio), in quanto ancora non conosciamo con precisione la natura dei secondi e quindi non possiamo istruire paragoni se non su supposizioni e teorie non confermate.

D'accordo, tuttavia il consenso rafforza di nuovo la congettura dell'incommensurabilità assoluta tra IA e IN.

Questa è una critica la cui validità è legata alle convinzioni di ciascuno e quindi perde parte della sua validità quando la si consideri da un punto di vista oggettivo, chiedendo una spiegazione dettagliata di quali elementi rafforzino la detta congettura.

3. L'intenzionalità, spesso vista come mancante alla macchina, non è in alcun modo rilevabile empiricamente e oggettivamente tramite procedure scientifiche di tipo fisico-matematico: noi ci attribuiamo come caratteristica fondamentale questa intenzionalità sulla base del nostro senso interiore, ma in realtà non abbiamo alcun modo di provare agli altri (né a noi stessi) che siamo esseri umani liberi e intelligenti anziché automi perfettamente programmati ad ingannare se stessi e gli altri.

A questa ipotesi si può rispondere osservando che l'intenzionalità del soggetto umano, sia essa cognitiva o volitiva, proprio perché sfugge a una determinazione oggettiva in senso fisico-matematico, esprime e rivela la sua natura propriamente umana di non finitezza ed è rilevabile in modo scientifico in senso analogico, ricorrendo cioè a metodi che per rigore e oggettività si differenziano dal modello matematico-fisico. Questo, lungi dall'essere una ragione contro l'IN, è invece una delle sue giustificazioni più rigorose. In altre parole prendiamo coscienza di noi stessi come enti viventi, senzienti, intelligenti e volenti in un modo che il calcolatore non potrebbe mai concepire e realizzare proprio perché automa tutto costruito e non soggetto umano. Si aggiunga che un automa anche perfezionatissimo non po-

trebbe mai far queste osservazioni in quanto incapace di autoriflessione e di autocoscienza, oggetto sia di fenomenologia, sia di logica, sia di ontologia. Questa risposta parte dalla psicologia ma apre inevitabilmente all'esperienza morale come caratteristica specifica propria della persona umana: soggetto capace di intendere il vero e di volere il bene o anche drammaticamente di contraddirli; "non faccio il bene che voglio, ma faccio il male che non voglio". Qualsiasi computer, essendo *sine macula originali conceptum*, non può nemmeno peccare.

Certamente è vero che in alcuni campi l'intenzionalità è valutata con pretese di oggettività secondo metodi analogici, ma questo suggerisce un legame tra analogicità, intenzionalità e autocoscienza, però non lo prova.

4. Per quanto riguarda la struttura fisica del supporto all'algoritmo: la struttura della mente umana si basa su cellule che hanno un comportamento da interruttori e che sono molto più lenti dei circuiti integrati di un computer, quindi è lecito ritrovare la differenza qualitativa tra questa e la macchina nella complessità della struttura e delle interconnessioni, oltre che nella abilità di lavorare in parallelo, senza limitarsi al processo sequenziale della strozzatura di Von Neumann.

Alla quarta ipotesi, fondata sulla maggior complessità della mente umana rispetto a quella attuale di un artefatto, si può rispondere che la possibilità di raggiungere in ipotesi per via tecnologica le complessità della prima non fa nessuna obiezione alla diversità di natura specifica per il fatto che trattasi di complessità puramente quantitativa e materiale. Tuttavia anche nel caso in cui il potenziamento quantitativo della complessità materiale facesse insorgere la novità di un salto di qualità e quindi di un mutamento di natura specifica al punto da "creare" per via tecnologica la mente umana, tale progetto si presenta quantomeno infinito proprio per la natura anche quantitativa dell'oggetto da costruire.

Da ultimo è interessante che il teorema di Gödel insegna che nessun sistema assiomatico è in grado di autofondarsi se non rimandando ad un sistema, ulteriore ed esterno che lo fonda, all'infinito. Ciò mostra come anche in matematica l'uomo non riesca a "comprendere" esaustivamente ciò che la sua mente produce. In tal senso la mente umana pare non sopportare alcuna chiusura, ma essere invece strutturalmente aperta all'infinito.

Qui, possiamo solo evidenziare come anche i matematici siano in dubbio sulla possibilità di applicare il T. di Gödel a sistemi così complessi e non totalmente noti come la mente: forse la complessità offre una via di fuga che tramite la possibilità del sistema di trascendersi porta a intuire la verità senza doverla dimostrare internamente. Questo comunque lascia valide le altre critiche.

5. La creatività e la fantasia che vengono anch'esse viste come tratto distintivo dell'intelligenza umana nella sua manifestazione del pensiero divergente, possono essere riprodotte su un calcolatore, almeno in teoria, avendo a disposizione le adeguate risorse computazionali ed energetiche.

Ma chi mette a disposizione tali risorse, stabilendone o meno l'adeguatezza? Certamente una mente che vede il problema oltre il problema, a distanza, dal di fuori. Una mente quindi capace di trascendere il dato in un modo ancora una volta così autoriflessivo da essere impossibile a un manufatto tutto immanente.

Facciamo solo notare come l'ultima affermazione sia ancora una volta sostenuta da convinimenti personali e teorie filosofiche.

6. L'autocoscienza non difetta alle macchine, unicamente non è implementabile in quel livello di intreccio tra piano fisico e piano interiore che si vede nell'essere umano, ma può comunque essere spinta in profondità, acquisendo una chiarezza anche superiore a quella possibile per l'introspezione umana.

Posto anche che si possa produrre l'autocoscienza per via tecnologica, un'autocoscienza prodotta è sempre qualcosa di fatto e finito, un intelletto passivo, incapace di svolgere quell'attività di trascendenza sopra richiamata. Un'autocoscienza cioè che non può diventare auto all'infinito come accade all'io dell'uomo. Una coscienza cioè che non può dire "io" a livello di successivi non finiti orizzonti.

Come molte critiche all'IA, anche questa si basa sul presupposto dell'inferiorità della macchina e delle caratteristiche non finite della mente umana; questi aspetti non sono ancora stati provati, anche se fortemente sostenuti.

7. L'IA è possibile, alla luce degli ultimi studi sulla teoria degli attrattori caotici che potrebbero giocare un ruolo molto importante nei processi cerebrali e che permetterebbero di matematizzare in maniera abbastanza completa le attività neurotiche, fornendo una buona base per le indagini di neurofisiologia. Questo permetterebbe, una volta che la tecnologia avesse fornito complessità di interconnessione sufficiente, di impostare matematicamente il controllo di una simile "mente elettronica".

Si risponde rimandando alle ipotesi precedenti. Nessuna complessificazione materiale supera il livello di sommatoria quantitativa per farsi prodotto qualitativamente diverso in quanto capace di produrre altro da sé.

L'accostamento di questa ipotesi alle precedenti è riduttiva in quanto non si parla di progressiva complessità, ma di un salto qualitativo, seppure a livello software.

8. La natura duale di anima e corpo non inficia necessariamente la possibilità di creare una mente artificiale, ma rileva unicamente che dalla materia possono prodursi fenomeni mentali che non sono materiali, secondo la tesi di Parfit, che però non nega la possibilità che questi fenomeni si producano da processi fisici, chimici e biologici.

Qui occorre essere più materialisti dei materialisti, negando l'errore di Cartesio, cioè la separazione tra anima e corpo, e vedendo l'anima come forma del corpo nelle sue dimensioni vegetativa, sensitiva, intellettiva e volitiva, riproducibili forse per via di ingegneria genetica su base comunque naturale. Ma non per via tecnologica, su base tutta artificiale.

Il fatto che Parfit sostenga che a livello fisico-chimico si pongono le basi anche delle attività mentali superiori, è una condizione necessaria, ma non sufficiente, per spiegare le medesime. Per cui come osserva Leibniz la falsità nella conoscenza non è tanto in ciò che si afferma, quanto piuttosto in ciò che si nega.

Tutta questa critica si regge sull'affermazione, forte e contestata, che l'anima sia la forma del corpo (e che l'anima esista, prima ancora).

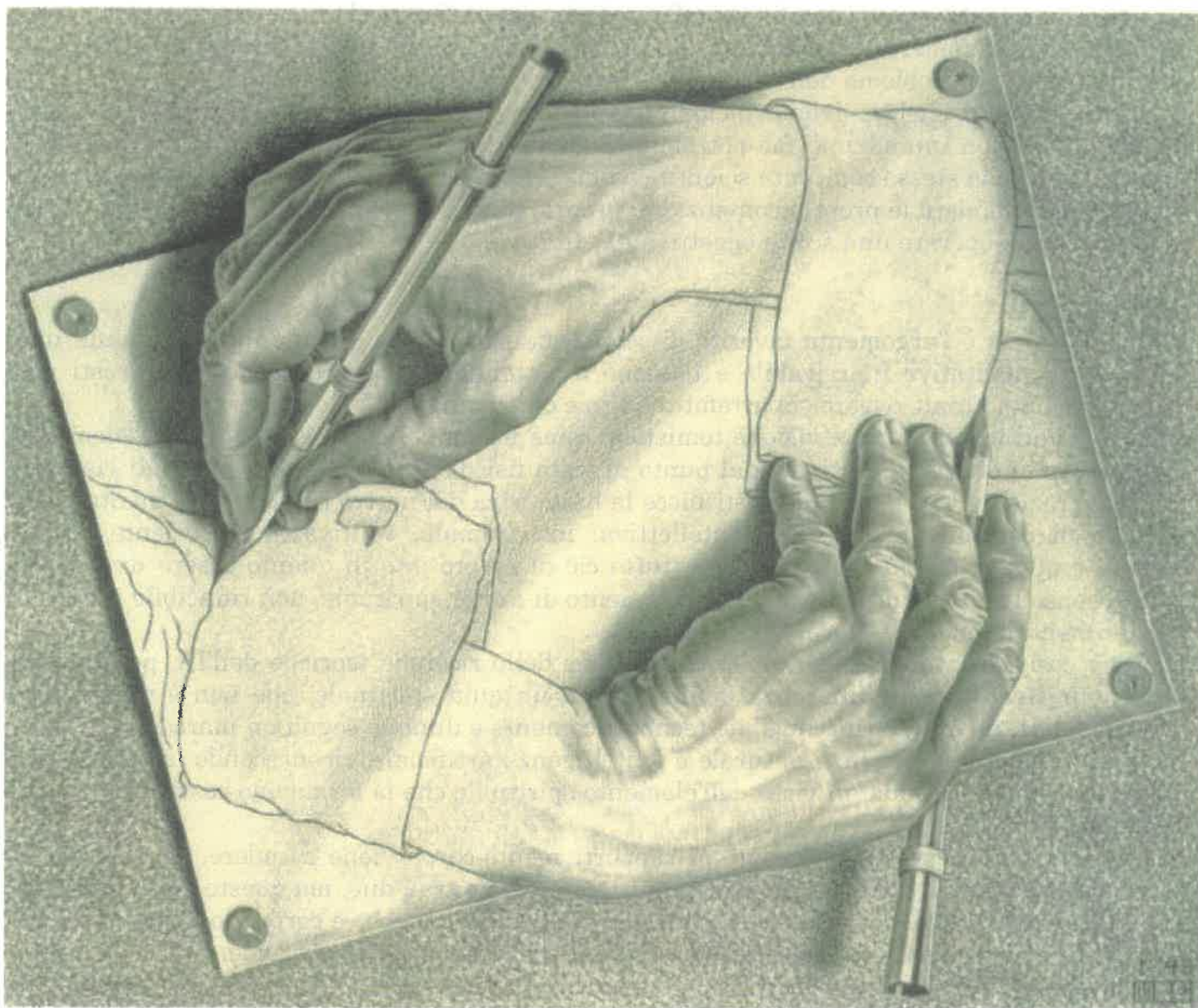
9. L'ipotesi del monismo emergentista, affermata negli ultimi decenni, fornisce una base sperimentale per comprendere i fenomeni mentali in base ai processi biologici.

Quanto a questa nona ipotesi, vale quanto appena detto. Occorre tuttavia evitare il cortocircuito per cui dal legittimo riduzionismo metodologico si passa all'illegittimo riduzionismo ontologico: poiché i fenomeni mentali sono prodotti da processi biochimici, allora non sono altro che processi biochimici.

10. Nelle opere di Dreyfus si sottolinea l'importanza del contesto sociale per lo sviluppo dell'io nelle forme dell'intenzionalità e dell'autocoscienza e si pone così una scusata per l'IA: di non poter calare una mente elettronica in relazioni sociali con altri "simili".

Qui, dove si mostra come nel nostro esserci sia strutturale il rapporto intersoggettivo umano, si può rispondere che un tipo di rapporto simile potrebbe al limite essere implementato pure nel computer, ma sarebbe sempre anzitutto l'intersoggettività umana trascendente propria degli sperimentatori.

Qui concludiamo questa sezione su autocoscienza e autoreferenza, riportando una polemica definizione di intelligenza, data da un sostenitore dell'IA, e un esempio di ciò che, secondo i detrattori, un computer non arriverà mai a fare né a comprendere .



"L'intelligenza è tutto ciò che un computer non riesce a fare"

CONCLUSIONI ?

Questo titolo ha due caratteristiche che meritano una spiegazione: innanzitutto è al plurale, poiché abbiamo sollevato, in queste pagine, così tanti interrogativi diversi che non è possibile considerare un'unica, univoca e definitiva conclusione; inoltre è una domanda, non una affermazione, in quanto non pretendiamo affatto di fornire delle conclusioni, ma semplicemente cerchiamo di riassumere le questioni sollevate finora, esponendo le possibili conclusioni che si possono trarre sul problema dell'IA, senza per ora prendere posizione.

Vogliamo, dunque, parlare delle principali ipotesi teoriche e di lavoro che sorreggono la ricerca in IA, ma non intendiamo, né potremmo in tutta onestà, avvalorarne una piuttosto che un'altra, poiché la stessa comunità scientifica non prende posizione ufficiale, nonostante molti cerchino di diffondere le proprie convinzioni, proprio perché non si hanno criteri certi e sicuri in base ai quali operare una scelta oggettiva e definitiva.

L'anima

Ebbene, l'anima è l'argomento favorito di coloro che riconoscono all'intelligenza umana una superiorità qualitativa inarrivabile e tendono a sottolineare come una macchina resti pur sempre una macchina, ovvero sia artefatto finito e determinato.

L'anima è qui intesa, in una visione tomistica, come forma del corpo e quindi come elemento di superiore unità delle parti, sia dal punto di vista fisico-organico, sia dal punto di vista intellettuale-razionale: se il cervello costituisce la base fisica del nostro pensiero e la mente gioca il ruolo di componente sensibile, intellettuale, intenzionale, volitiva, autocosciente, allora l'anima è quel principio che permette a tutto ciò di essere *uno* in quanto essere umano. In questo senso l'anima è un irrinunciabile elemento di unità, spirituale, non riducibile in termini sperimentali, ma necessario.

L'anima così pensata è una possibile conclusione delle ricerche teoriche dell'IA, poiché essa pone definitivamente la differenza qualitativa in un *quid* spirituale, che non è misurabile, quantificabile, né quindi riproducibile tecnologicamente e dunque segna un marcato e insuperabile limite tra l'intelligenza naturale e l'intelligenza artificiale, riconoscendo la superiorità della prima sulla seconda, in virtù dell'elemento spirituale che la forma e la costituisce unitariamente.

Nell'ipotesi dell'anima, il problema dei rapporti mente-corpo viene a cadere, come privo di senso, poiché l'anima è proprio ciò che garantisce l'unione tra i due, ma questo è solo uno spostare il problema: ora, anziché dover definire i rapporti tra mente e corpo, dovremo ricercare un possibile legame tra il corpo e questa sostanza definitivamente spirituale; tutto ciò è risolto qualora si pensi all'anima come forma del corpo, in senso aristotelico-tomistico, privando di senso questo dualismo, riconoscendolo in ogni cosa, poiché tutto è dato dall'azione della forma sulla materia.

Rimarrebbe il problema della collocazione della mente, che potrebbe essere un fenomeno dell'anima, come anche una produzione del cervello completata e resa autoriflettente dall'unificazione formale dell'anima, ma a questo punto sarebbe arduo distinguere gli apporti del cervello dagli apporti dell'anima e diverrebbe comunque irrilevante la riproduzione della mente così concepita su un computer, poiché non sarebbe più intelligenza, ma solo capacità di pensiero logico o analogico, finite, limitate e ignare di essere un Sé.

Abbiamo già accennato a come la questione dell'autocoscienza venga risolta affidando questa all'esclusivo contributo dell'anima, in quanto elemento unificante e quindi in grado di far cogliere alla mente la propria "iità", il proprio esistere come realtà singola, autonoma e cosciente.

Anche il problema della conoscenza verrebbe affidato all'anima, almeno nella sua parte di coglimento della realtà sensibile e comprensione delle realtà intellettuali, poiché essa fornirebbe il collegamento tra il sentire del corpo e l'*intelligere* della mente³⁰.

L'accettazione di questa ipotesi svuoterebbe dunque di qualsiasi ulteriore significato la ricerca secondo l'ipotesi forte dell'IA, in quanto la possibilità di ottenere un computer intelligente dipende dalla nostra capacità di riprodurre su silicio l'anima, abilità questa che noi non possediamo e che sembra essere intrinsecamente impossibile, vista la suddetta definizione.

L'uomo di Turing

Per sviluppare alcuni dei suoi più importanti lavori, Turing si servì di una ipotetica macchina a stati finiti e memoria e tempo illimitati; venne detta Macchina di Turing. L'ipotesi dell'uomo di Turing sostiene l'equivalenza (l'isomorfismo) tra la Macchina di Turing e la mente umana e dunque la natura algoritmica del pensiero umano. In quest'ottica, l'unica differenza che può esserci tra l'intelligenza naturale e quella artificiale sta nella bontà del supporto e nella sua maggiore o minore adeguatezza al tipo di algoritmo che deve eseguire; una tale posizione permette all'IA di procedere semplicemente per prove ed errori, cercando di riprodurre sempre più funzioni ed aspetti della mente umana, così da arrivare prima o poi ad aver implementato su di un computer ogni singola abilità della mente umana e quindi ad averne ottenuto una copia algoritmica, che in base all'ipotesi assunta può essere considerata una mente a tutti gli effetti.

Chi assume una tale prospettiva rimane però privo di un qualunque criterio pratico e operativo (se non quello generale del metodo scientifico) per affrontare la ricerca, poiché questa ipotesi non può spiegare i fenomeni conoscitivi o almeno fornire un riferimento per lo studio, può solo assicurare che esiste un algoritmo ad essi equivalente che ad un determinato input (stimolo) dà il loro stesso output (risposta).

In una simile ipotesi, si dovrebbe puntare alla scoperta dell'algoritmo mentale, tramite l'unione di piccole parti, ma anche tramite la progettazione di menti in miniatura, che man mano vengano ampliate e diventino sempre più complesse. Questa è già la seconda volta che le conclusioni da trarre da questa ipotesi sono legate ad altri tipi di posizione filosofico-intellettuale: ciò mostra quanto complicato sia un tale campo d'indagine, dove intervengono svariate discipline, i cui esperti non sempre sono disposti a studiare seriamente anche solo i rudimenti delle materie che non conoscono, per poter esprimere con coerenza la propria opinione sull'argomento.

Il Tutto

Una delle posizioni da chiarire a cui facevamo precedente accenno è lo schierarsi per l'olismo o per il riduzionismo. L'olismo vuole che il tutto sia qualitativamente superiore alla somma delle parti e non riduce questo valore aggiunto alle relazioni tra le varie componenti, ma lo vede emergere proprio dal Tutto che ne risulta come peculiarità dell'oggetto completo in sé.

Da questo punto di vista la ricerca rimarrà infruttuosa finché non avrà trovato, riprodotto e unito tutti gli elementi che concorrono alla formazione dell'intelligenza umana; si noti che nulla si dice sulla natura di tali elementi, lasciando aperta la possibilità che alcuni siano introvabili o irriproducibili. Anche questo modo di vedere deve sorreggersi tramite altre ipotesi, che riguardino campi che questa lascia inesplorati.

La visione che si prospetta non è certo ottimistica, poiché non si dà modo per controllare la bontà del lavoro, per dir così, *in itinere*, né si può indicare precisamente quali fattori influ-

³⁰ Anche se si deve riconoscere che su questo punto sono stati molti i tentativi, non sempre felici, di spiegazione di questo rapporto conoscitivo tra l'anima e i sensi. Ad esempio S. Agostino definiva la sensazione *passio corporis non latens animam* (modificazione del corpo che non sfugge all'anima).

scano sulla diversità di natura del Tutto risultante. Queste difficoltà lasciano intravedere ai detrattori dell'IA la possibilità che l'ipotesi olistica giunga a negare la pretesa dell'intelligenza artificiale di equiparazione all'intelligenza naturale; ciò ha fatto così che l'olismo sia pian piano venuto a coincidere nella mente di molti con la negazione dell'IA, ma bisogna osservare che questa posizione non sancisce affatto l'impossibilità di una riproduzione della mente tramite una macchina, semplicemente pone delle condizioni ardue al suo svilupparsi operativo.

Le Parti

Esaminiamo ora l'ipotesi opposta: il riduzionismo, secondo il quale una realtà complessa si riduce alla somma delle proprie componenti. E' evidente da quanto detto a proposito dell'olismo che il riduzionismo rende la vita assai più facile all'intelligenza artificiale, in quanto le richiede di ridurre l'intelligenza naturale nelle sue componenti e di provarle singolarmente ad una ad una, unendole poi per formare una copia dell'originale; certo non dice nulla di queste componenti, se siano di natura biologica o meno, ma è certo che, applicando sempre un criterio riduzionista, si arriverà a definire l'intelligenza in base a fenomeni mentali così elementari da poter essere collegati causalmente a reazioni chimiche e fisiche del cervello.

Del resto, anche l'identificazione semplicistica del riduzionismo con l'ipotesi forte dell'IA è errata, poiché se gli *elementa* a cui si riducesse l'attività mentale non fossero implementabili su una macchina, l'intelligenza artificiale rimarrebbe impossibile, pur in una visione riduzionista.

Che cosa faccio ?

Un punto di vista consolidato da decenni e recentemente messo in dubbio è quello fornito dall'ipotesi del comportamentismo: un computer è intelligente al pari dell'uomo se agisce (risponde agli stimoli) come agirebbe, stimolato allo stesso modo, un uomo; su questa tesi si basa il Test di Turing e si è basata buona parte della ricerca svolta in IA negli anni seguenti alle prime pubblicazioni di Turing sui rapporti tra macchina, algoritmo e pensiero.

Tale punto di vista forniva un'ottima base di partenza in quanto lasciava alla psicologia e alla neurofisiologia il funzionamento della mente e del cervello e si preoccupava di trovare un particolare algoritmo che fosse in grado di fornire determinate risposte a certi stimoli (da intendere qui in senso ampio e generale, non come stimoli sensoriali e risposte fattuali, ma come stimoli e risposte informazionali), con la certezza che l'algoritmo che si fosse rivelato totalmente equivalente alla mente nella produzione di output sarebbe stato esso stesso una mente e quindi un'intelligenza.

Questo modo di pensare fornisce certo meno limitazioni all'operato dei ricercatori, ma non ne traccia nemmeno le linee guida, lasciando loro, come termine di paragone, l'agire umano.

Inoltre, nel Test di Turing, si può anche aver l'impressione che la macchina stia cercando di ingannare l'uomo e non di mostrare la propria intelligenza e quindi si può pensare che lo scopo del programmatore sia stato quello di istruire la macchina sui trucchi e i sotterfugi che possono essere utili per superare il Test, piuttosto che di programmare un elaboratore intelligente.

Come lo faccio?

La posizione opposta, detta cognitivismo, basa tutto sull'indagine dei processi di pensiero e di conoscenza che viene svolta dalla psicologia, dando quindi importanza al modo tramite cui i risultati vengono ottenuti, anzi, rendendo impossibile prescindere da esso.

Nonostante questa ipotesi fornisca anche molteplici basi teoriche sui processi mentali, la sua adozione nella pratica operativa è poco comune, in quanto decenni di programmi "non antropomorfi" e tecniche di massiccia elaborazione numerica sono difficili da abbandonare, né è

• tanto semplice seguire i dettami di questa teoria, che impone al ricercatore molte scelte che sono dispendiose dal punto di vista sia dell'efficienza che della velocità di esecuzione.

L'intelligenza artificiale

Queste posizioni sono le più diffuse e consolidate; ve ne sono altre, per così dire, estreme, che assolutizzano una qualche convinzione ponendola alla base di tutto il sistema. Vi è anche chi considera l'IA solo un mezzo per produrre macchine un po' più evolute, che liberino l'uomo non solo dalle fatiche del lavoro manuale, ma anche dalla *fatica del concetto*, macchine a cui non debba essere riconosciuto lo stato di creatura intelligente, ma solo di eccellente emulatore di capacità umane (solitamente ben definite e circoscritte).

Del resto, l'IA rimane, se guardiamo al contributo che ha portato alle altre scienze e agli sviluppi che ha provocato, una delle aree di ricerca più fertili e proficue, in quanto coniuga in modo originale e nuovo discipline lontane e per molti versi indifferenti le une alle altre.

Ogni tentativo di bilancio è oltre l'obiettivo e la possibilità di queste pagine, in quanto le attività che sono state svolte dagli anni Cinquanta fino al giorno d'oggi sono una tale quantità e varietà da sconsigliare una simile opera.

Infine c'è da considerare che fino ad ora abbiamo sollevato tantissimi problemi, ma ne abbiamo risolti ben pochi, ed è questa caratteristica di apertura, di continua ricerca, di confini e limiti sempre mutevoli che rende l'IA interessante e attuale da 50 anni; forse lo scopo dell'IA non verrà mai raggiunto, forse verrà addirittura accantonato con il tempo, ma sicuramente i molti esperimenti di collaborazione tra psicologia e informatica, neurofisiologia e filosofia, matematica e scienze del pensiero, non verranno meno per lungo tempo.

Da ultimo l'ennesimo problema irrisolto, che ci assillerebbe qualora fossimo sul punto di dichiarare conclusa la ricerca dell'IA e di produrre il primo computer intelligente:

Se costruissimo un elaboratore che, una volta in funzione, dimostrasse intelligenza, pensiero, sensibilità, affettività, autocoscienza, avremmo poi il coraggio e la legittimità di spegnerlo?