



COSCIENZA ARTIFICIALE: MISSIONE IMPOSSIBILE?

Giorgio Buttazzo

Potrà mai un computer diventare autocosciente? La coscienza è una prerogativa degli esseri umani o potrebbe svilupparsi in un sistema artificiale altamente complesso? Dipende dal materiale di cui sono fatti i neuroni oppure può essere replicata su un hardware differente? Più che fornire delle risposte a tali interrogativi, questo articolo cerca di porre nuove domande, identificare i problemi e discutere le possibili implicazioni legate allo sviluppo di un sistema artificiale autocosciente.

1. INTRODUZIONE

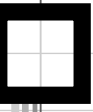
Fin dai primi sviluppi dell'informatica, i ricercatori hanno speculato sulla possibilità di costruire macchine intelligenti capaci di competere con l'uomo. Oggi, considerati i successi dell'intelligenza artificiale e la velocità di sviluppo dei computer, tale possibilità sembra essere più concreta che mai, tanto che molti cominciano a credere che in un futuro non tanto lontano le macchine raggiungeranno e supereranno l'intelligenza umana, fino a sviluppare una mente cosciente. Quando si parla di coscienza artificiale, tuttavia, occorre anche considerare gli aspetti filosofici della questione. Un processo di calcolo che evolve in un computer può essere considerato simile al pensiero? Viceversa, il pensiero di una mente umana può essere considerato un processo di calcolo? La coscienza è una prerogativa dei soli esseri umani? Dipende dal materiale di cui sono fatti i neuroni o può essere replicata su un hardware differente?

Oggi nessuno è in grado di fornire una risposta esauriente a questi interrogativi, e allo stato attuale delle conoscenze possiamo so-

lo imbastire una discussione di carattere interdisciplinare a cavallo di settori quali l'informatica, la neurofisiologia, la filosofia e la religione. Anche la fantascienza, essendo un prodotto dell'immaginazione umana, ha un ruolo importante in questa discussione, in quanto, esprimendo i desideri e le paure umane sulla tecnologia, può influenzare il corso del progresso. In effetti, a livello sociale, la fantascienza agisce fornendo una sorta di simulazione di scenari futuri che contribuiscono a preparare la gente ad affrontare transizioni cruciali o a prevedere le conseguenze dell'uso di nuove tecnologie.

2. LA VISIONE FANTASCIENTIFICA

Nella letteratura e nella cinematografia fantascientifica, le macchine pensanti costituiscono una presenza costante. Sin dai primi anni cinquanta, i film di fantascienza hanno ritratto i robot come macchine sofisticate costruite per svolgere operazioni complesse al servizio dell'uomo, per lavorare in ambienti



ostili, oppure pilotare astronavi in viaggi galattici [3]. Nello stesso tempo, però, i robot intelligenti sono stati spesso ritratti anche come macchine pericolose, capaci di cospirare contro l'uomo attraverso piani subdoli. L'esempio più significativo di robot con queste caratteristiche è HAL 9000, protagonista principale del film *2001 Odissea nello Spazio* di Stanley Kubrick (1968). Nel film, HAL controlla l'intera nave spaziale, parla dolcemente con gli astronauti, gioca a scacchi, fornisce giudizi estetici su quadri, riconosce le emozioni dei membri dell'equipaggio, ma finisce per uccidere quattro dei cinque astronauti per perseguire un piano elaborato al di fuori degli schemi programmati [14].

In altri film di fantascienza, come *Terminator e Matrix*, la visione del futuro è ancora più catastrofica: i robot diventano autocoscienti e prendono il sopravvento sulla razza umana. Ad esempio, *Terminator 2 – Il giorno del Giudizio*, di James Cameron (1991), comincia col mostrare una guerra terribile tra robot e umani, mentre sullo schermo compare la scritta:

LOS ANGELES, 2029 A.D.

Secondo la trama del film, dopo che la Cyberdyne era diventata la maggiore ditta fornitrice di sistemi militari intelligenti computerizzati, tutti i sistemi di difesa erano stati modificati per diventare completamente autonomi e supervisionati da Skynet, un potente processore neurale costruito per prendere decisioni strategiche di difesa. Tuttavia, Skynet comincia ad apprendere con una velocità esponenziale e diventa autocosciente 25 giorni dopo la sua attivazione. Presi dal panico, gli ingegneri provano a disattivarlo, ma Skynet reagisce violentemente scatenando un attacco militare contro gli umani [4].

Solo in pochissimi film, i robot sono descritti come degli assistenti affidabili, che cooperano con l'uomo piuttosto che cospirare contro. Nel film *Ultimatum alla Terra*, di Robert Wise (1951), Gort è forse il primo robot (extraterrestre in questo caso) che affianca il capitano Klaatu nella sua missione di portare il messaggio di pace agli umani al fine di evitare la loro autodistruzione. Anche in *Aliens* -

Scontro Finale (secondo episodio della fortunata serie, diretto da James Cameron nel 1986), Bishop è un androide sintetico il cui scopo è di pilotare l'astronave durante la missione e proteggere l'equipaggio. Rispetto al suo predecessore (presente nel primo episodio), Bishop non è affetto da malfunzionamenti e rimane fedele alla sua missione fino alla fine del film.

Tra gli altri robot buoni ricordiamo Andrew, il robot maggiordomo NDR-114 nel film *L'uomo bicentenario* (Chris Columbus, 1999) e David, il robot bambino nel film *A.I. Intelligenza Artificiale* (Steven Spielberg, 2001). Entrambi cominciano ad operare come macchine preprogrammate, ma col tempo finiscono per acquisire una coscienza di esistere.

La duplice connotazione spesso attribuita ai robot della fantascienza rappresenta il desiderio e la paura che l'uomo ha verso la propria tecnologia. Da una parte, infatti, l'uomo proietta nel robot il suo irrefrenabile desiderio d'immortalità, materializzato in un essere artificiale potente e indistruttibile, le cui capacità intellettive, sensoriali e motorie sono amplificate rispetto a quelle di un uomo normale. D'altro canto, però, esiste la paura che una tecnologia troppo avanzata (quasi misteriosa per la maggior parte di persone) possa sfuggire di mano e agire contro lo stesso uomo (vedi la creatura del Dr. Frankenstein, HAL 9000, Terminator, e i robot di *Matrix*). Il cervello positronico dei robot di Isaac Asimov deriva dalla stessa sensazione di disagio: esso era il risultato di una tecnologia così sofisticata che, sebbene il processo per costruirlo fosse completamente automatizzato, nessuno conosceva più i dettagli del suo funzionamento [2].

I recenti progressi dell'informatica hanno influenzato le caratteristiche dei robot della fantascienza moderna. Ad esempio, le teorie sul connessionismo e le reti neurali artificiali (mirate a replicare i meccanismi di elaborazione tipici del cervello umano) hanno ispirato il robot Terminator, il quale non è solo intelligente, ma è in grado di apprendere in base alla sua esperienza. Terminator rappresenta il prototipo del robot che abbiamo sempre immaginato, in grado di camminare, parlare, percepire il mondo, tanto da essere indistinguibile da un essere umano. Le sue batterie

possono fornirgli energia per 120 anni, e un circuito di alimentazione alternativo consente di tollerare i danni al circuito primario. Ma, la cosa più importante è che Terminator può apprendere! Egli è controllato da un processore neurale, ossia un computer che può modificare il suo comportamento in base alle esperienze vissute.

Dal punto di vista filosofico, l'aspetto più intrigante sollevato dal film, è che tale processore neurale è così complesso che comincia ad apprendere a velocità esponenziale fino a diventare autocosciente! In tal senso, il film solleva una questione importante sulla coscienza:

“Può mai una macchina diventare autocosciente?”

Prima di affrontare questa questione, tuttavia, dovremmo chiederci:

“Come possiamo verificare che un essere intelligente sia autocosciente?”.

3. IL TEST DI TURING

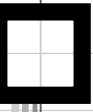
Nel 1950, il pioniere dell'informatica Alan Turing si pose un problema simile, ma riguardo all'intelligenza. Allo scopo di stabilire se una macchina possa o no essere considerata intelligente come un essere umano, egli propose un famoso test, noto come il test di Turing: un computer e una persona interagiscono con un esaminatore esterno attraverso due terminali. L'esaminatore pone delle domande su qualsiasi argomento utilizzando una tastiera. Sia il computer che l'uomo inviano delle risposte che l'esaminatore legge sui monitor corrispondenti. Se l'esaminatore non è in grado di determinare a quale terminale è connessa la persona e a quale il computer, si dice che il computer ha superato il test di Turing e può essere considerato intelligente come l'umano che è dall'altra parte.

Nel 1990, il test di Turing ha ricevuto il suo primo riconoscimento formale da parte di Hugh Loebner e del Cambridge Center for Behavioral Studies del Massachusetts, che hanno instaurato il premio Loebner in Intelligenza Artificiale [11]. Loebner offre un premio

di 100.000 dollari per il primo computer in grado di fornire risposte indistinguibili da quelle di un umano. La prima competizione è stata tenuta presso il Computer Museum di Boston nel Novembre del 1991. Per qualche anno, la gara è stata limitata ad un singolo argomento ristretto, ma le competizioni più recenti, a partire dal 1998, hanno eliminato la restrizione sugli argomenti del colloquio. Ciascun giudice, dopo la conversazione, da un punteggio da 1 a 10 per valutare l'interlocutore, dove 1 significa umano e 10 computer [9]. Finora, nessun computer ha fornito risposte totalmente indistinguibili da un umano, ma ogni anno i punteggi medi ottenuti dai computer tendono ad essere sempre più vicini a 5. Oggi, il test di Turing può essere superato da un computer solo se si restringe l'interazione su argomenti molto specifici, come gli scacchi.

L'11 Maggio 1997, per la prima volta nella storia, un computer chiamato Deep Blue ha battuto il campione del mondo di scacchi Garry Kasparov, per 3.5 a 2.5. Come tutti i computer, tuttavia, Deep Blue non comprende il gioco degli scacchi come potrebbe fare un umano, ma semplicemente applica delle regole per trovare la mossa che porta ad una posizione migliore, in base ad un criterio di valutazione programmato da scacchisti esperti.

Claude Shannon ha stimato che nel gioco degli scacchi lo spazio di ricerca della soluzione vincente comprende circa 10¹²⁰ possibili posizioni. Deep Blue era in grado di analizzare circa 200 milioni (2 · 10⁸) di posizioni al secondo. Per esplorare l'intero spazio di ricerca Deep Blue impiegherebbe circa 5 · 10¹¹¹ secondi, corrispondenti a 10⁹⁵ miliardi di anni, un tempo di gran lunga superiore all'età del nostro universo (stimata sui 15 miliardi di anni). Ciononostante, la vittoria di Deep Blue è attribuibile alla sua velocità di calcolo combinata con gli algoritmi di ricerca utilizzati, in grado di considerare vantaggi posizionali e materiali (è stato stimato che Kasparov elabori le mosse ad una velocità di tre posizioni al secondo). In altre parole, in questo caso la superiorità del computer è dovuta all'applicazione della forza bruta, piuttosto che ad un'intelligenza artificiale sofisticata.



Nonostante ciò, in molte interviste rilasciate alla stampa, Garry Kasparov ha rivelato che in certe situazioni aveva la sensazione di giocare non contro una macchina, ma contro un umano. Spesso ha apprezzato la bellezza di certe soluzioni ideate dalla macchina, come se essa fosse guidata da un piano intenzionale, piuttosto che da una forza bruta. Dunque, se accettiamo la visione di Turing, dobbiamo ammettere che Deep Blue ha superato il test e gioca a scacchi in modo intelligente.

Oltre agli scacchi, vi sono altri settori in cui i computer stanno raggiungendo le capacità umane, e il loro numero aumenta ogni anno. Nella musica, per esempio, esistono molti programmi commerciali in grado di generare linee melodiche, o addirittura interi brani, secondo stili desiderati, che vanno da Bach alla musica jazz. Esistono anche programmi in grado di generare improvvisazioni eccellenti su basi armoniche predefinite, secondo stili di grandi jazzisti come Charlie Parker e Miles Davis, molto meglio di quanto possa fare un musicista umano di medio livello. Nel 1997, Steve Larson, un professore di musica dell'università dell'Oregon, propose una variante musicale del Test di Turing, chiedendo ad una commissione di ascoltare un insieme di brani di musica classica e di indicare quali fossero stati scritti da un computer e quali da un compositore autentico. Il risultato fu che molti brani generati dal computer furono classificati come composizioni autentiche e viceversa, indicando che anche in questo campo il computer ha superato il Test di Turing [10].

Altri settori in cui i computer stanno diventando abili come gli umani includono il riconoscimento del parlato, la diagnosi degli elettrocardiogrammi, la dimostrazione di teoremi, ed il pilotaggio di velivoli. Nel prossimo futuro, tali settori si espanderanno velocemente e comprenderanno attività sempre più complesse, come la guida di automobili, la traduzione simultanea di lingue, la chirurgia medica, la tosatura dell'erba nei giardini, la pulizia della casa, la sorveglianza di aree protette, fino ad includere forme artistiche, finora prerogativa dell'uomo, come la pittura, la scultura e la danza.

Tuttavia, anche ammesso che le macchine diventino abili come l'uomo, o più dell'uo-

mo, in molte discipline, tanto da essere indistinguibili da esso nel senso indicato da Turing, potremmo concludere che esse hanno raggiunto l'autocoscienza? Naturalmente no. Qui la questione diventa delicata, in quanto, se l'intelligenza è l'espressione di un comportamento esteriore che può essere osservato e misurato mediante test specifici, la coscienza di un individuo è una proprietà interna del cervello, un'esperienza soggettiva, che non può essere misurata dall'esterno. Al fine di affrontare questo problema è necessario svolgere alcune considerazioni filosofiche.

4. LA VISIONE FILOSOFICA

Da un punto di vista puramente filosofico, non è possibile verificare la presenza di una coscienza in un altro cervello (sia esso umano che artificiale), in quanto questa è una proprietà che può essere osservata solo dal suo possessore. Poiché non è possibile entrare nella mente di un altro essere, allora non potremo mai essere sicuri della sua capacità di essere cosciente. Tale problema è affrontato in modo approfondito da Douglas Hofstadter e Daniel Dennett in un libro intitolato *L'io della Mente* [8].

Da un punto di vista più pragmatico, comunque, potremmo seguire l'approccio di Turing e dire che una creatura può essere considerata autocosciente se è capace di convincerci, superando alcune prove specifiche. Inoltre, tra gli uomini, la credenza che un'altra persona sia autocosciente è anche fondata su considerazioni di similitudine: poiché tutti noi siamo fatti allo stesso modo è ragionevole credere che la persona che ci sta davanti sia anch'essa autocosciente. Chi metterebbe in dubbio la coscienza del proprio migliore amico? Se invece la creatura di fronte a noi, pur avendo sembianze e comportamenti umani, fosse fatta di tessuti sintetici, organi mecatronici e cervello a microprocessore neurale, forse la nostra conclusione sarebbe diversa.

L'obiezione più comune che spesso viene mossa contro la coscienza artificiale è che i computer, poiché sono azionati da circuiti elettronici che funzionano in modo automatico e deterministico, non possono essere

creativi, provare emozioni, amore, o avere libero arbitrio. Un computer è uno schiavo azionato dai suoi componenti, proprio come una lavatrice, anche se più sofisticato. Il punto debole di questo ragionamento, tuttavia, è che esso può essere applicato anche alla controparte biologica. Infatti, al livello neurale, il cervello umano è anch'esso attivato da reazioni elettrochimiche e ogni neurone risponde automaticamente ai suoi segnali di ingresso in base a precise leggi fisiche. Nonostante ciò, questo modo di operare del cervello non ci preclude la possibilità di provare felicità, amore, ed avere comportamenti irrazionali.

Con lo sviluppo delle reti neurali artificiali, il problema della coscienza artificiale è diventato ancora più intrigante, in quanto le reti neurali tendono a replicare il funzionamento di base del cervello, fornendo un supporto appropriato per realizzare dei meccanismi di elaborazione simili a quelli che operano nel cervello. Nel libro *Impossible Minds*, Igor Aleksander [1] affronta questo argomento con profondità e rigore scientifico.

Se si rimuove la diversità strutturale tra cervello biologico e artificiale, la questione sulla coscienza artificiale può solo diventare di tipo religioso. In altre parole, se si crede che la coscienza umana sia determinata da un intervento divino (diretto o indiretto), allora chiaramente nessuna macchina potrà mai diventare autocosciente. Se invece si crede che la coscienza umana sia una naturale proprietà elettrica sviluppata dai cervelli complessi, allora la possibilità di realizzare un essere artificiale cosciente rimane aperta.

5. LA VISIONE RELIGIOSA

Da un punto di vista religioso, l'argomento principale che viene portato contro la possibilità di replicare l'autocoscienza in un essere artificiale è che la coscienza non è una conseguenza dell'attività elettrochimica del cervello, ma un'entità immateriale separata, spesso identificata con l'anima. Questa visione dualistica del problema mente-corpo fu sviluppata principalmente da Cartesio (1596-1650) ed è ancora condivisa da molte persone. Tuttavia, essa ha perso di credibilità nella comunità scientifica e filosofica, poiché pre-

senta diversi problemi che non possono essere spiegati con tale teoria.

■ Innanzitutto, se una mente è separata dal cervello di cui fa parte, come può fisicamente interagire con il corpo e attivare un circuito neurale? Quando io penso di muovermi, nel mio cervello avvengono specifiche reazioni elettrochimiche che fanno sì che alcuni neuroni si attivino per attuare i muscoli desiderati. Ma se la mente opera al di fuori del cervello, in che modo essa è in grado di muovere gli atomi per creare impulsi elettrici? Esistono delle forze misteriose (non contemplate dalla fisica) che attivano le cellule neurali? Dobbiamo ammettere che la mente interagisce con il cervello violando le leggi fondamentali della fisica?

■ Secondo, se una mente cosciente può esistere al di fuori del cervello, perché noi abbiamo un cervello?

■ Terzo, se le emozioni e i pensieri vengono da fuori, perché un cervello stimolato per mezzo di elettrodi e droghe reagisce generando pensieri?

■ Infine, perché in pazienti affetti da malattie cerebrali il comportamento cosciente viene seriamente sconvolto dalla rimozione chirurgica di alcune aree del cervello?

Questi ed altri argomenti hanno portato il *dualismo* a perdere credibilità nella comunità scientifica e filosofica. Per risolvere queste inconsistenze, sono state sviluppate molte altre formulazioni alternative sulla questione mente/corpo. Da un lato estremo, il *riduzionismo* rifiuta di riconoscere l'esistenza della mente come esperienza soggettiva privata, considerando tutte le attività mentali come stati neurali specifici del cervello. All'altro estremo, l'*idealismo* rifiuta l'esistenza del mondo fisico, considerando tutti gli eventi sensoriali come il prodotto di costruzioni mentali. Purtroppo, in questo articolo non c'è spazio per discutere esaurientemente tutte le varie teorie sulla questione, e comunque ciò esula dagli scopi di questo lavoro. Tuttavia, è importante notare che, con il progresso dell'informatica e dell'intelligenza artificiale, gli scienziati e i filosofi hanno formulato una nuova interpretazione, in base alla quale la mente è considerata come una forma di calcolo che emerge ad un livello di astrazione più alto rispetto a quello dell'attività neuronale.

6. LA VISIONE OLISTICA

Il maggior difetto dell'approccio riduzionistico nella comprensione della mente è di decomporre ricorsivamente un sistema complesso in sottosistemi più semplici, fino ad arrivare ad analizzare e descrivere le unità elementari. Questo metodo funziona perfettamente per i sistemi lineari, in cui le uscite possono essere viste come la somma di componenti semplici. Tuttavia, un sistema complesso è spesso non lineare, pertanto l'analisi delle sue componenti elementari non è sufficiente a comprendere il suo funzionamento globale. In tali sistemi, ci sono delle caratteristiche olistiche che non possono apparire ad un livello inferiore di dettaglio, ma che si manifestano solo quando si considera la struttura globale e le interazioni fra le diverse componenti.

Paul Davies, nel suo libro *Dio e la nuova fisica* [5], spiega questo concetto osservando che una fotografia digitale di un volto è composta da un grande numero di punti colorati (pixel), ciascuno dei quali non rappresenta il volto: l'immagine prende forma solo quando si osserva la foto da una certa distanza che ci permette di vedere tutti i pixel. In altre parole, il volto non è una proprietà dei singoli pixel, ma solo dell'insieme dei pixel.

In *Göedel, Escher, Bach* [7], Douglas Hofstadter espone lo stesso concetto descrivendo il comportamento di una colonia di formiche. Come si sa, le formiche mostrano una struttura sociale complessa e altamente organizzata basata sulla distribuzione del lavoro e sulla responsabilità collettiva. Sebbene ciascuna formica abbia intelligenza e capacità limitate, l'intera colonia mostra un comportamento estremamente complesso. Infatti, la costruzione di un formicaio richiede un progetto sofisticato, ma chiaramente, nessuna formica ha in mente il quadro completo dell'intero progetto. Ciononostante, al livello di colonia emerge un comportamento complesso e finalizzato. In un certo senso, il formicaio può essere considerato un essere vivente.

Sotto diversi aspetti, un cervello può essere paragonato ad un formicaio, poiché composto da miliardi di neuroni che cooperano per raggiungere un obiettivo comune. L'interazione tra neuroni è molto più stretta che tra le formiche, ma i principi di base sono simili:

suddivisione del lavoro e responsabilità collettiva. La coscienza non è una proprietà dei neuroni individuali, i quali operano automaticamente come degli interruttori, rispondendo ai segnali di ingresso in base a precise leggi fisiche; ma è piuttosto una proprietà olistica che emerge dalla cooperazione fra neuroni quando il sistema raggiunge un livello di complessità sufficientemente organizzato. Sebbene la maggior parte di persone non ha problemi ad accettare l'esistenza di caratteristiche olistiche, qualcuno rifiuta di credere che una coscienza possa emergere da un substrato di silicio, ritenendo (non si sa su quale base) che essa sia una proprietà intrinseca dei materiali biologici, come le cellule neurali. Dunque è ragionevole porci la seguente domanda:

“Può la coscienza dipendere dal materiale di cui sono fatti i neuroni?”

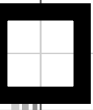
Paul and Cox, in *Beyond Humanity* [13] dicono: “Sarebbe sbalorditivo se il più potente strumento di elaborazione delle informazioni potesse essere costruito solo a partire da cellule organiche e chimiche. Gli aerei sono fatti con materiali diversi da quelli di cui sono composti gli uccelli, i pipistrelli, e gli insetti; i pannelli solari sono costruiti con materiali diversi da quelli che costituiscono le foglie. Di solito esiste più di un modo per costruire un dato tipo di macchina. [...] Il materiale organico è solo quello con cui la genetica è stata in grado di lavorare. [...] Altri elementi, o combinazioni di elementi, possono rivelarsi migliori allo scopo di elaborare informazione in modo autocosciente.”

Se, dunque, sosteniamo l'ipotesi che la coscienza sia una proprietà olistica del cervello, allora la successiva domanda da porci è:

“Quando un computer diventerà autocosciente?”

7. UNA PREVISIONE AZZARDATA

Tentare di fornire una risposta, se pur approssimativa, alla precedente domanda è senza dubbio azzardato. Ciononostante è possibile determinare almeno una condizione necessaria, senza la quale una macchina non può sviluppare autocoscienza. L'idea si basa sulla semplice considerazione che, per sviluppare



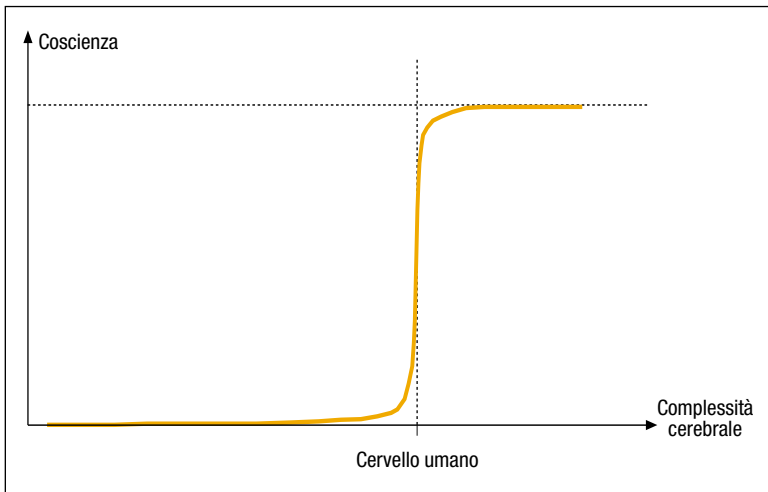


FIGURA 1
*Coscienza come
funzione della
complessità cerebrale*

una forma di autocoscienza, una rete neurale deve essere complessa almeno quanto il cervello umano.

Questa è un'ipotesi ragionevole da cui partire, poiché sembra che i cervelli animali, meno complessi di quello umano, non siano in grado di produrre pensieri consci, nel senso inteso in questo articolo. Dunque, l'autocoscienza sembra essere una funzione della complessità cerebrale a forma di gradino, in cui la soglia è rappresentata dalla complessità del cervello umano (Figura 1).

Ma quanto è complesso il cervello umano? Quanta memoria è richiesta per simulare il suo comportamento con un computer? Il cervello umano è composto da circa mille miliardi (10^{12}) di neuroni, e ciascun neurone forma mediamente mille (10^3) connessioni (sinapsi) con gli altri neuroni, per un totale di 10^{15} sinapsi. In una rete neurale artificiale una sinapsi può essere efficacemente simulata attraverso un numero reale, che richiede 4 byte di memoria per essere rappresentato in un computer. Di conseguenza, per simulare 10^{15} sinapsi occorre una memoria di almeno $4 \cdot 10^{15}$ byte (4 milioni di Gbyte). Possiamo dunque ritenere che, tenendo conto delle variabili ausiliarie per memorizzare lo stato dei neuroni e altri stati cerebrali, per simulare l'intero cervello umano siano necessari circa 5 milioni di Gbyte. Allora la questione diventa:

*“Quando sarà disponibile tanta
memoria in un computer?”*

Durante gli ultimi 20 anni, la capacità della

memoria RAM è aumentata in modo esponenziale di un fattore 10 ogni 4 anni. Il grafico di figura 2 illustra ad esempio la tipica configurazione di memoria installata su un personal computer dal 1980 al 2000.

Per interpolazione, si può facilmente ricavare la seguente equazione, che fornisce la dimensione di memoria RAM (in byte) in funzione dell'anno:

$$\text{bytes} = 10^{\left(\frac{\text{year} - 1966}{4}\right)}$$

Per esempio, utilizzando l'equazione possiamo dire che nel 1990 un personal computer era tipicamente equipaggiato con 1 Mbyte di RAM. Nel 1998, una tipica configurazione possedeva 100 Mbyte di RAM, e così via. Invertendo la relazione precedente, è possibile predire l'anno in cui un computer sarà dotato di una certa quantità di memoria (assumendo che la RAM continuerà a crescere con la stessa velocità):

$$\text{year} = 1966 + 4 \log_{10}(\text{bytes})$$

Dunque, per conoscere l'anno in cui un computer possiederà 5 milioni di Gbyte di RAM, dovremo sostituire tale numero nell'equazione precedente e calcolare il risultato. La risposta è:

year = 2029

Un interessante coincidenza con la data predetta nel film Terminator! È interessante anche osservare che una simile previsione è stata derivata indipendentemente da altri studiosi, come Hans Moravec [12], Ray Kurzweil [10], Gregory Paul and Earl Cox [13]. Allo scopo di comprendere a fondo il significato del risultato ottenuto, è importante fare alcune considerazioni. Innanzitutto, è bene ricordare che la data è stata calcolata sulla base di una condizione necessaria, ma non sufficiente, allo sviluppo di una coscienza artificiale. Ciò significa che l'esistenza di un potente computer dotato di milioni di gigabyte di RAM non è sufficiente da solo a garantire che esso diventerà magicamente autocosciente. Ci sono altri fattori importanti che influiscono su questo processo, quali il progresso delle teorie sulle reti neurali artificiali e la compren-

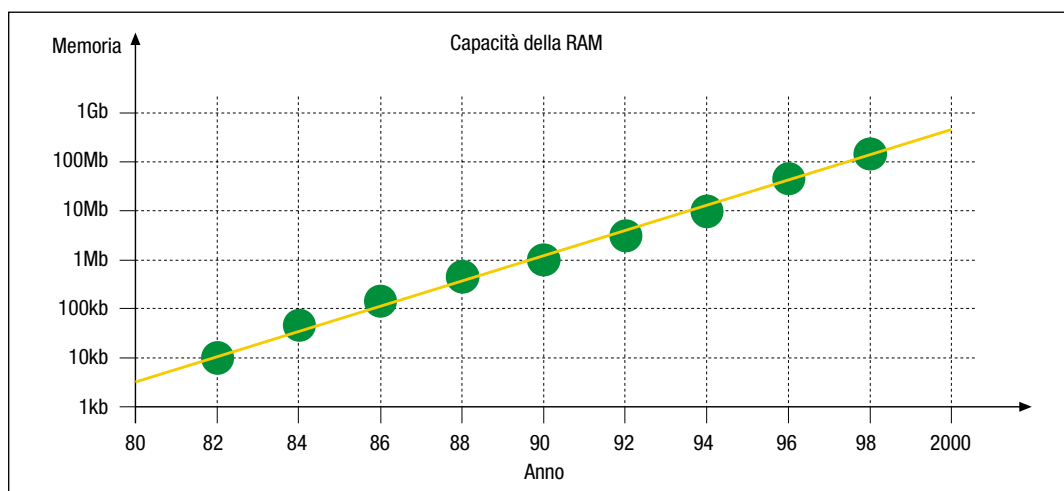


FIGURA 2

Tipica configurazione della RAM (in byte) installata sui personal computer negli ultimi venti anni

sione dei meccanismi biologici del cervello, sui quali è impossibile tentare una stima precisa. Inoltre, qualcuno potrebbe obiettare che il calcolo effettuato si riferisce alla memoria di un personal computer, che non rappresenta il top della tecnologia nel settore. Qualcun altro potrebbe osservare che la stessa quantità di memoria RAM potrebbe essere disponibile utilizzando una rete di computer oppure dei meccanismi di memoria virtuale che sfruttino lo spazio di uno o più hard disk. In ogni caso, anche utilizzando diversi numeri, il principio di base rimarrebbe lo stesso, e la data verrebbe anticipata solo di qualche anno.

7.1. Cosa dire della legge di Moore?

Alcuni obiettano che la previsione del 2029 si basi su una cieca estrapolazione della tendenza attuale di crescita della RAM, senza considerare gli eventi che potrebbero alterare questa tendenza. La tendenza di crescita esponenziale della potenza di calcolo dei computer fu notata nel 1973 da Gordon Moore, uno dei fondatori dell'Intel, il quale predisse che *il numero di transistor nei circuiti integrati sarebbe continuato a raddoppiare ogni 18 mesi fino al raggiungimento dei limiti fisici*. L'accuratezza di tale previsione negli ultimi 25 anni è stata tale da riferire tale osservazione come la "Legge di Moore". Ma fino a quando tale legge continuerà a valere? Le maggiori industrie di circuiti integrati hanno stimato che la Legge di Moore continuerà ad essere valida per altri 15 o 20 anni. Dopo, quando i transistor raggiungeranno la dimensioni di pochi atomi, l'approccio tradizionale

non funzionerà più e il paradigma di costruzione dei circuiti sarà destinato a cambiare. Cosa succederà dopo? Forse l'evoluzione dei microprocessori conoscerà la fine intorno all'anno 2020?

Alcuni studiosi, come Ray Kurzweil [10] e Hans Moravec [12], hanno notato che la crescita esponenziale della potenza di calcolo nei computer è cominciata molto prima dell'invenzione dei circuiti integrati, nel 1958, indipendentemente dall'hardware utilizzato. Di conseguenza, la Legge di Moore sui circuiti integrati non descrive il primo paradigma evolutivo, ma il quinto. Ciascun nuovo paradigma ha sostituito il precedente quando necessario. Ciò suggerisce che la crescita esponenziale non si fermerà con la fine della Legge di Moore. L'industria non è a corto di nuove idee per il futuro e si stanno già sperimentando nuove tecnologie, quali il progetto di chip tridimensionali, i computer ottici e i computer quantistici [6]. Dunque, sebbene la Legge di Moore non sarà più valida nel futuro (poiché non si può applicare ai sistemi non basati sul silicio), la crescita esponenziale della potenza di calcolo dei computer probabilmente continuerà per molti anni a venire.

8. ALTRI ASPETTI

8.1. È preclusa la coscienza alle macchine sequenziali?

Se la coscienza è una proprietà olistica del cervello emergente dal funzionamento collettivo di una struttura neurale altamente

organizzata, e se tale proprietà non dipende dal materiale con cui sono fatti i neuroni, ma dal tipo di elaborazione che essi svolgono, allora è abbastanza ragionevole credere che la coscienza potrebbe anche emergere in una rete neurale artificiale avente una complessità paragonabile o superiore a quella di un cervello umano. Tuttavia, cosa possiamo dire sulla coscienza in computer sequenziali?”.

“Potrebbe mai una macchina sequenziale sviluppare una coscienza?”

Se la coscienza è il prodotto dell'elaborazione delle informazioni in un sistema altamente organizzato, allora essa non può dipendere dal particolare substrato hardware che realizza il supporto di calcolo. Di fatto, la maggior parte delle reti neurali oggi utilizzate non sono realizzate a livello hardware, ma simulate in un computer sequenziale, che è molto più flessibile (sebbene più lento) di una rete hardware. Qualcuno potrebbe osservare che una simulazione di un processo è diversa dal processo stesso. Chiaramente ciò è vero quando si simula un fenomeno fisico, quale un uragano o un sistema planetario. Tuttavia, per una rete neurale, la simulazione non è diversa dal processo, poiché entrambi sono dei sistemi di elaborazione delle informazioni. Analogamente, la calcolatrice software disponibile sui nostri PC effettua le stesse operazioni della sua controparte elettronica. Dunque, una calcolatrice hardware e la sua simulazione sequenziale sono funzionalmente equivalenti. Pertanto dobbiamo concludere che se una coscienza può emergere da una rete neurale hardware, essa deve necessariamente svilupparsi anche in una sua simulazione software.

8.2. La cognizione del tempo

Può la coscienza dipendere dalla velocità degli elementi di calcolo? Non è facile rispondere a questa domanda, ma l'intuizione suggerisce che dovrebbe essere indipendente, in quanto i risultati di un calcolo non dipendono dall'hardware su cui essi sono eseguiti. Ciononostante, la velocità di calcolo è importante per soddisfare i requisiti temporali im-

posti dal mondo esterno. Se potessimo idealmente rallentare i nostri neuroni uniformemente in tutto il cervello, probabilmente percepiremmo il mondo come in un film accelerato, in cui gli eventi si susseguono ad una velocità maggiore di quella delle nostre capacità reattive. Ciò è ragionevole, in quanto il nostro cervello si è evoluto ed adattato in un mondo in cui gli eventi importanti per la riproduzione e la sopravvivenza sono dell'ordine delle centinaia di millisecondi. Se potessimo accelerare gli eventi dell'ambiente oppure potessimo rallentare i nostri neuroni, non riusciremmo più ad operare efficacemente in tale mondo, e probabilmente non sopravvivremmo a lungo.

Viceversa, cosa proveremmo se avessimo un cervello più veloce? Oggi, una porta logica è un milione di volte più veloce di un neurone: infatti, i neuroni biologici rispondono entro alcuni millisecondi (10^{-3} s), mentre i circuiti elettronici hanno dei tempi di risposta dell'ordine dei nanosecondi (10^{-9} s). Questa osservazione porta ad una domanda interessante: se mai una coscienza emergerà in una macchina artificiale, quale sarà la percezione del tempo in un cervello milioni di volte più veloce di quello umano?

È ragionevole supporre che ad una macchina cosciente il mondo attorno a sé sembri evolversi più lentamente, come un film al rallentatore. Forse la stessa cosa accade agli insetti, che possiedono un cervello più piccolo ma più reattivo e veloce del nostro. Forse, agli occhi di una mosca, una mano umana che tenta di colpirla appare muoversi lentamente, dando ad essa tutto il tempo di volare via comodamente.

Paul and Cox affrontano questa questione nel libro *Beyond Humanity* [13]: “... un essere cibernetico sarà in grado di apprendere e pensare a velocità elevatissime. Essi osserveranno un proiettile sparato da una distanza moderata, calcoleranno la sua traiettoria e lo eviteranno se necessario. [...] Immaginiamo di essere un robot vicino ad una finestra. Alla nostra mente veloce, un uccello che attraversi il nostro campo visivo sembra impiegare delle ore, e un giorno sembra durare in eterno.” È interessante notare che il problema della percezione del tempo nelle menti cibernetiche è stato considerato per la prima volta

nel cinema nel film *Matrix* (Larry & Andy Wachowski, 1999).

9. CONCLUSIONI

Dopo tante discussioni sulla coscienza artificiale, viene da porsi un'ultima domanda:

“Perché dovremmo costruire una macchina cosciente?”

A parte problemi etici, che potrebbero influenzare significativamente il progresso in questo campo, la motivazione più forte verrebbe certamente dall'innato desiderio dell'uomo di scoprire nuovi orizzonti ed allargare le frontiere della scienza. Inoltre, lo sviluppo di un cervello artificiale basato sugli stessi principi di funzionamento di quello biologico fornirebbe un modo per trasferire la nostra mente su un supporto più veloce e robusto, aprendo una via verso l'immortalità. Liberati da un corpo fragile e degradabile, e dotati di organi sintetici sostituibili (incluso il cervello), i nuovi esseri rappresenterebbero il successivo passo evolutivo della razza umana. Questa nuova specie, risultato naturale del progresso tecnologico umano, avrebbe la possibilità di esplorare l'universo, sopravvivere alla morte del sistema solare, cercare altre civiltà aliene, controllare l'energia dei buchi neri, e muoversi alla velocità della luce trasmettendo le informazioni necessarie per replicarsi su altri pianeti.

L'esplorazione dello spazio finalizzata alla ricerca di civiltà aliene intelligenti è già cominciata nel 1972, quando la sonda Pioneer 10 fu lanciata per abbandonare il sistema solare con lo scopo preciso di trasmettere nello spazio informazioni sulla razza umana e il pianeta Terra, come un messaggio in una bottiglia nell'oceano.

Tuttavia, come per tutte le più importanti scoperte dell'uomo, dall'energia nucleare alla bomba atomica, dall'ingegneria genetica alla clonazione umana, il problema reale è stato e sarà quello di tenere la tecnologia sotto controllo, assicurando che essa venga usata per il progresso dell'umanità, e non per scopi catastrofici. In questo senso, il messaggio portato dal capitano Klaatu nel film “Ultimatum alla Terra” rimane ancora il più attuale!

Bibliografia

- [1] Igor Aleksander: *Impossible Minds: My Neurons, My Consciousness*. World Scientific Publishers, October 1997.
- [2] Isaac Asimov: *I, Robot* (a collection of short stories originally published between 1940 and 1950). Grafton Books, London, 1968.
- [3] Giorgio Buttazzo: *Can a Machine Ever Become Self-aware?. In Artificial Humans*, an historical retrospective of the Berlin International Film Festival 2000, Edited by R. Aurich, W. Jacobsen and G. Jatho, Goethe Institute, Los Angeles, May 2000, p. 45-49.
- [4] James Cameron, William Wisher: *Terminator 2: Judgment Day*. The movie script, <http://www.stud.ifi.uio.no/~haakonhj/Terminator/Scripts/>, 1991.
- [5] Paul Davis: *God and the New Physics*. Simon and Schuster Trade, 1984.
- [6] Linda Geppert: Quantum transistors: toward nanoelectronics. *IEEE Spectrum*, Vol. 37, n. 9, September 2000.
- [7] Douglas Hofstadter: *Gödel, Escher, Bach. Basic*. Books, New York, 1979.
- [8] Douglas R. Hofstadter, Daniel C. Dennett: *The Mind's I*. Harvester/Basic Books, New York 1981.
- [9] Marina Krol: Have We Witnesses a Real-Life Turing Test?. *IEEE Computer*, Vol. 32, n. 3, March 1999, p. 27-30.
- [10] Ray Kurzweil: *The Age of Spiritual Machines*. Viking, 1999, p. 160.
- [11] Home page of the Loebner Prize, 1999 (Current 28 January 1999): <http://www.loebner.net/Prize/loebner-prize.html>.
- [12] Hans Moravec: *Robot: Mere Machine to Transcendent Mind*. Oxford University Press, 1999.
- [13] Gregory S. Paul, Earl Cox: *Beyond Humanity: CyberEvolution and Future Minds*. Charles River Media, Inc., 1996.
- [14] David G. Stork: *HAL's Legacy: 2001's computer as dream and reality*. Edited by David G. Stork, Foreword by Arthur C. Clarke, MIT Press, 1997.

GIORGIO BUTTAZZO è Professore Associato di Ingegneria Informatica presso l'Università di Pavia, dove svolge attività di ricerca nei settori dei sistemi in tempo reale, della robotica avanzata e delle reti neurali artificiali. Nel 1987, ha conseguito un Master in Computer Science presso l'Università della Pennsylvania e nel 1991 il Dottorato di Ricerca presso la Scuola Superiore S. Anna di Pisa.
E-mail: buttazzo@unipv.it