

# INTELLIGENZA E COSCIENZA

## L'IA tra Searle e Dennett.

### Sviluppi dell'Intelligenza Artificiale

Matteo Fini – Paola Milani

## Introduzione

Questo capitolo è essenzialmente suddiviso in tre parti. La prima presenta due posizioni filosofiche antitetiche, ciascuna delle quali conduce ad una diversità di approccio, e quindi di risoluzione, ai problemi sollevati dall'Intelligenza Artificiale. Da una parte, troviamo il funzionalista<sup>1</sup> Daniel C. Dennett, il quale, partendo da una particolare concezione dell'intenzionalità e della coscienza<sup>2</sup>, giunge ad un'adesione e ad un'apologia dei programmi sostenuti dall'IA; dall'altra, si muove John R. Searle, con la teoria del "naturalismo biologico"<sup>3</sup> e la convinzione che gli stati mentali siano solo parzialmente riducibili a stati fisici.

La seconda parte del capitolo mostra invece l'evoluzione degli studi nel campo dell'IA, dedicando particolare attenzione ai più recenti sviluppi, agli scopi della ricerca ed alle aspettative per il futuro.

In quest'ottica, il capitolo si chiude con una disamina delle più recenti tendenze nell'ambito dello studio della mente, sottolineando come si stia rendendo necessario - per giungere davvero alla creazione di una 'macchina che pensa'- affiancare all'Intelligenza Artificiale una nuova disciplina: la Coscienza Artificiale.

---

<sup>1</sup> In filosofia, con il termine "funzionalismo" si intende la dottrina secondo cui gli eventi o stati mentali (credenza, desiderio, volizione, ecc.) sarebbero qualificati da funzioni, ossia da ruoli operazionali, anziché da una specifica costituzione materiale. In altre parole, "gli stati mentali sono individuati dalle loro relazioni funzionali con gli altri stati mentali, con gli input sensoriali e con gli output comportamentali" (cfr. M. Di Francesco, *Introduzione alla filosofia della mente*, Carocci, Roma 2002, p.94).

Il funzionalismo fa parte di una più ampia corrente, definita "materialismo", che cerca di spiegare i processi mentali deprivandoli della componente soggettiva ("qualia") e riducendoli a processi fisici. E' importante, in questa sede, accennare anche al "funzionalismo computazionale", che confluisce poi nell'IA. Tale teoria aggiunge al funzionalismo l'idea che la mente non è altro che un programma per calcolatore, e che il pensiero consiste solo nel manipolare simboli, inseriti in una rete funzionale che può essere spiegata in termini computazionali.

<sup>2</sup> In generale, con il termine di "intenzionalità" si intende il riferirsi di un atto o stato mentale ad un determinato oggetto intenzionale, ovvero la connessione che l'atto o lo stato ha con un certo oggetto, indipendentemente dal sussistere o meno di questo oggetto nella realtà esterna. In altre parole, l'"essere intenzionali" è una proprietà di alcuni stati mentali di essere *diretti verso*, o *relativi a*, individui o stati di cose. Questa caratteristica, la direzionalità, indica dunque che tali stati mentali hanno un *contenuto*, cioè verso cui si direzionano. Si possono quindi "definire intenzionali gli stati mentali con un contenuto" (cfr. S. Gozzano, *Storia e teorie dell'intenzionalità*, Laterza, Bari 1997, p. VIII).

Per esempio, se si crede che domani pioverà, lo stato mentale di credenza è diretto verso il fatto che domani pioverà, e quest'ultimo è il contenuto (significato) dello stato intenzionale.

Il concetto di coscienza, nel senso di consapevolezza di sé e degli oggetti ai quali essa si rivolge, è strettamente legato a quello di intenzionalità. A partire da Husserl e dalla fenomenologia, infatti, l'aspetto per cui la coscienza è sempre 'coscienza di qualcosa', ossia ha necessariamente un oggetto quale termine di riferimento, è definito come la sua 'intenzionalità'.

<sup>3</sup> E' la teoria, introdotta da Searle, secondo cui la coscienza, e i fenomeni ad essa legati (intenzionalità, soggettività, intelligenza, ecc.), è un fenomeno biologico come qualsiasi altro.

Esso si basa su due assunti:

- la coscienza, compresa la sua soggettività, è causata da processi nel cervello;
- gli stati mentali coscienti (intenzionali e non) sono essi stessi caratteristiche di livello superiore del cervello.

Ecco perché Searle, come vedremo nei prossimi paragrafi, sostiene che 'la mente sia causata dal cervello'.

## Searle e Dennett: due differenti approcci all'IA

E' noto, all'interno del dibattito contemporaneo legato all'Intelligenza Artificiale, il contenzioso tra Dennett e Searle, entrambi impegnati da anni in ricerche legate al funzionamento della mente.

Le divergenze tra Searle e Dennett in questo settore rappresentano, tuttavia, come vedremo, l'esito più ovvio di una profonda frattura teorica, connessa alla questione di cosa significhi effettivamente 'avere una mente'<sup>4</sup>. Infatti, i due filosofi abbracciano delle concezioni di mente, intenzionalità, coscienza molto diverse l'una dall'altra, e soprattutto concepiscono in modo dissimile la natura dei cosiddetti "poteri causali"<sup>5</sup> del cervello: questo disaccordo di fondo non poteva che generare due visioni opposte sulle possibilità di successo dell'intelligenza artificiale.

Buona parte delle osservazioni di Dennett prende avvio dall'analisi del famoso "argomento della stanza cinese", addotto da Searle a sostegno della sua teoria dell'irriducibilità del mentale al fisico. Questo argomento, riportato nelle righe seguenti, è stato pubblicato nel 1980 sulle pagine di *Behavioral and Brain Sciences*:

"Immaginate che un uomo che non conosce il cinese sia chiuso a chiave in una stanza e abbia a disposizione una serie di ideogrammi e un programma per calcolatore in grado di rispondere a quesiti formulati in quella lingua. Gli input del sistema così costruito sarebbero costituiti da domande formulate in cinese, mentre gli output sarebbero risposte nella stessa lingua. Possiamo supporre che il programma a disposizione dell'uomo sia di tale qualità da far sì che le sue risposte siano indistinguibili da quelle che fornirebbe un parlante madrelingua. Ciò nonostante né la persona chiusa dentro la stanza, né alcuna parte del sistema comprende veramente il cinese; inoltre, visto che il calcolatore così programmato non ha alcuna proprietà addizionale rispetto al sistema nel suo complesso, esso stesso, in quanto calcolatore, non comprende affatto il cinese"<sup>6</sup>.

Tramite questo esperimento mentale, Searle si proponeva di mostrare come la mente sia dotata di una *semantica* (quindi di un significato, di una comprensione) che un programma per calcolatore non potrà mai acquisire, in quanto dotato solo di *sintassi*, e di sottolineare quindi l'impossibilità di paragonare una mente umana al funzionamento di un calcolatore<sup>7</sup>.

---

<sup>4</sup> E' il punto cruciale della trattazione. In generale, per Searle, 'avere una mente' significa possedere una coscienza, ossia stati mentali non riducibili ai processi neurofisiologici del cervello (da cui tuttavia sono causati) perché dotati di soggettività ontologica: gli stati coscienti esistono, cioè, solo in quanto esperiti da un agente. E' il problema dei "qualia" (stati soggettivi dell'esperienza): secondo Searle, essi esistono e non possono essere ridotti a fenomeni oggettivamente quantificabili, come gli stati funzionali di un sistema fisico, o gli stati computazionali del cervello.

Posizione, quest'ultima, assunta invece da Dennett, il quale nega completamente l'esistenza dei "qualia", e identifica la coscienza in "una serie di programmi per computer implementati nel cervello" (cfr. D. C. Dennett, *Coscienza. Che cos'è*, tr. it. Rizzoli, Milano 1992, cit. in E. Carli, *Cervelli che parlano*, B. Mondadori, Milano 1997, p.67). Il problema della coscienza, insomma, per Dennett riguarda solamente gli aspetti meccanici e tecnologici del cervello e della mente.

<sup>5</sup> I poteri causali del cervello, oggetto della disputa, consisterebbero nella capacità di quest'organo di originare gli stati mentali descritti precedentemente. Secondo Searle, il cervello "causa" la mente, nel senso che produce eventi mentali dotati di intenzionalità; il problema è che non è chiaro in che modo questo avvenga e in cosa consistano tali poteri. Per Dennett, invece, questo tipo di poteri causali non esiste, perché non esistono nemmeno stati interni misteriosi e ineffabili; gli unici poteri causali che esistono sono quelli che regolano le relazioni di input e output complesse e i comportamenti (verbali e non verbali) con cui si manifestano.

<sup>6</sup> John R. Searle, *La riscoperta della mente*, tr. it. Bollati Boringhieri, Torino 1994, p.61.

<sup>7</sup> In breve, secondo Searle, la macchina è dotata solo di una sintassi, cioè esegue operazioni di calcolo su elementi specificati per via formale, mentre la mente umana possiede anche una semantica, ossia è in grado di comprendere quello che sta facendo. Anche la mente umana manipola simboli, e in questo può essere simile ad una macchina, ma in più è in grado di "produrre significati" per questi simboli, cosa che il calcolatore non può fare. Potrebbe riuscirci solo se fosse in grado di riprodurre i poteri causali del cervello, che, nella teoria searlina, causano gli stati mentali.

Vediamo ora quale critica avanzi Dennett contro l'argomento della stanza cinese di Searle e quali siano i punti chiave della sua analisi del rapporto mente-macchina.

### **D. C. Dennett: una posizione a favore dell'IA**

Riprendiamo schematicamente le conclusioni cui giungeva Searle al termine dell'esposizione del suo 'argomento della stanza cinese'.

L'argomentazione searlina si articola in quattro premesse e quattro conclusioni:

- i cervelli causano le menti, nel senso che la produzione di stati mentali dotati di intenzionalità compete esclusivamente ai particolari poteri causali propri del cervello umano organico;
- la sintassi non è sufficiente per la semantica;
- i programmi da calcolatore sono interamente definiti attraverso la loro struttura formale, o sintattica;
- le menti hanno contenuti mentali; in particolare hanno contenuti semantici;
- CONCLUSIONE 1→nessun programma di calcolatore è di per sé sufficiente a dare una mente ad un sistema. In breve, i programmi non sono menti, e non sono di per sé sufficienti per avere una mente.
- CONCLUSIONE 2→il modo in cui le funzioni del cervello causano la mente non può consistere nella semplice esecuzione di programmi da calcolatore;
- CONCLUSIONE 3→qualsiasi altra cosa che fosse in grado di causare la mente dovrebbe avere poteri causali almeno equivalenti a quelli del cervello.
- CONCLUSIONE 4→per qualsiasi artefatto che noi potessimo costruire che avesse stati mentali equivalenti agli stati mentali umani, l'implementazione di un programma da calcolatore sarebbe di per sé insufficiente. Piuttosto, l'artefatto dovrebbe avere poteri equivalenti ai poteri del cervello umano<sup>8</sup>.

Quindi, l'idea che Searle difende è che di fatto nessun programma di un computer di per sé potrebbe mai essere sufficiente per produrre fenomeni mentali aventi un contenuto intenzionale.

La posizione sostenuta da Dennett può sembrare apparentemente simile, ma in realtà si pone su un piano radicalmente diverso; egli cerca proprio di dimostrare che non c'è possibilità di equivoco, se si conferisce ad alcune nozioni il giusto significato.

Dennett afferma che, allo stato attuale, "non c'è alcun modo in cui un computer digitale elettronico potrebbe esser programmato, così da essere in grado di produrre ciò che un cervello organico umano, con i suoi particolari poteri causali, può dimostrativamente produrre: il controllo dell'attività intenzionale rapida, intelligente, esibita da esseri umani normali"<sup>9</sup>.

Come si vede, le due posizioni finiscono, ad una prima lettura, col sembrare in gran parte la stessa cosa. In realtà, come emergerà dopo un'analisi più approfondita, si tratta di due visioni nettamente distinte.

Il punto cruciale delle due proposizioni si impenna su una differente concezione di "poteri causali". Searle considera l'esistenza di due livelli di poteri causali, quelli che presiedono alla produzione di stati mentali e quelli che originano (insieme con il resto del sistema nervoso) relazioni input-output.

Queste ultime sono ciò che la macchina può imitare dell'uomo, e si manifestano nel comportamento verbale e non verbale dell'individuo. Il fatto, tuttavia, che esteriormente una macchina si comporti al pari di un essere umano non significa che posseda i medesimi stati intenzionali, proprio perché ad un programma mancherà sempre una

<sup>8</sup> Cfr. J. R. Searle, *Mente, cervello, intelligenza*, tr. it. Bompiani, Milano 1987, pp. 30-32.

<sup>9</sup> D. C. Dennett, *L'atteggiamento intenzionale*, tr. it. Il Mulino, Bologna 1992, pp. 432-433.

Ricordiamo che per Dennett l'intenzionalità è una strategia che permette di prevedere e spiegare il comportamento di un sistema. In questo senso, non è una qualità intrinseca degli individui, ma solo un primo livello di spiegazione che viene presupposto e attribuito in una prima fase e che viene scomposto in sottofunzioni più semplici prive di intelligenza o intenzionalità. Per un ulteriore approfondimento della teoria dell'intenzionalità, si rimanda al paragrafo successivo.

semantica, ossia la capacità di produrre significati. La macchina, in qualche modo, non è 'consapevole' di essere quello che è o di quello che sta facendo (esattamente come l'individuo nella stanza non sa il cinese pur riuscendo a parlare in cinese), e questo perché l'unico tipo di causalità che essa può attuare è quella determinata da relazioni di causa-effetto, per cui ad un certo stimolo corrisponde una certa risposta.

I poteri causali hanno, per Dennett, invece, tutt'altro tipo di valenza. Egli critica Searle in merito alla sua teoria della causazione intenzionale<sup>10</sup> proprio per rimarcare una delle idee di fondo della sua visione: la distinzione avanzata da Searle tra poteri causali, in realtà, non ha senso di essere.

La misteriosità del concetto di 'poteri causali', così come è proposta da Searle, ha spesso offerto il fianco alle critiche dei suoi avversari, in ragione del fatto che sembrava rappresentare il punto debole della teoria.

Dennett è ancora più radicale rispetto ad altri critici: egli sostiene, infatti, che il livello della causazione intenzionale non esiste. I poteri causali ammessi da Dennett si riferiscono solamente ai poteri causali che occorrono per guidare attraverso il mondo reale un corpo capace di muoversi da un luogo all'altro.

Quindi, per Dennett, se un corpo, un'entità si comporta in un modo tale da esibire una certa intelligenza, è sufficiente per dire che è dotato di una mente, al pari dell'uomo. Questa posizione ha fatto sì che Searle lo definisse un "comportamentista"; in realtà Dennett va oltre, in quanto non si limita a ridurre gli stati mentali a puri comportamenti osservabili, ma arriva ad asserire che non c'è nulla di rilevante *oltre* tali comportamenti. E' per questo che, quando l'IA sarà riuscita a comprendere appieno il funzionamento del cervello umano organico, cioè il modo in cui esso produce poteri causali che conducono il corpo attraverso l'esistenza, il sentire, il vedere, il parlare, ecc., e sarà in grado di riprodurlo, allora il risultato sarà necessariamente un'entità dotata di una mente (nell'unico senso in cui, secondo Dennett, si può avere una mente).

Resta da capire come Dennett risolve il problema di quale sia l'elemento che caratterizza la mente umana e che determina quel controllo dell'attività intenzionale definibile come "intelligente".

Ne "L'atteggiamento intenzionale"<sup>11</sup>, Dennett afferma che " la velocità fa parte dell' 'essenza' dell'intelligenza. Se non si possono calcolare le parti attinenti dell'ambiente in trasformazione abbastanza rapidamente per provvedere a se stessi, non si è *praticamente* intelligenti, per quanto complessi si sia".

La struttura del cervello umano possiede un certo grado di velocità a trattare gli input e gli output come si presentano: è questa caratteristica che determina il comportamento del soggetto e che ne regola l'attività intenzionale intelligente.

Quindi, la teoria di Dennett è che ciò che una macchina dovrebbe imitare del cervello per poter dire di essa che possiede una mente non è qualche misterioso e oscuro potere causale che ad un livello superiore produce stati mentali, bensì la giusta velocità d'elaborazione delle informazioni.

Nelle parole di Dennett: "Consideriamo due diverse implementazioni dello stesso programma (...). In un certo senso, entrambe le implementazioni hanno le stesse capacità - entrambe 'computano la stessa funzione' - ma in virtù di nient'altro se non della sua velocità, una di esse avrà poteri causali di cui l'altra è priva(...)" e ancora: "(...) la velocità relativa è cruciale nel permettere che occorran i tipi esatti di sequenze di interazione ambiente - organismo"<sup>12</sup>.

---

<sup>10</sup> La "causazione intenzionale", per Searle, indica la connessione che vige tra la capacità di rappresentazione della mente e le relazioni causali con il mondo. Essa è essenziale per il funzionamento dell'intenzionalità e per la sopravvivenza dell'uomo nel mondo.

Quindi, nella causazione intenzionale, lo stato intenzionale è causa dello stato di cose che esso rappresenta o lo stato di cose che esso rappresenta è causa dello stato intenzionale.

<sup>11</sup>D. C Dennett, *L'atteggiamento intenzionale*, cit., p. 435.

<sup>12</sup>Ivi, p. 443.

Ancora non si conosce esattamente come funzioni il cervello<sup>13</sup>, e questo, secondo Dennett, è l'unico vero impedimento che ci separa dalla realizzazione di una macchina pensante. Ciò che Dennett vuole sottolineare è che può essere vero che un computer digitale non potrà mai attivare il programma giusto abbastanza velocemente da riprodurre in tempo reale il funzionamento del cervello, cioè quei poteri causali attraverso cui esso produce il controllo della rapida, intelligente, attività intenzionale. Tuttora non si è ancora riusciti in questo intento e, anzi, nonostante i successi conseguiti si è lontani dall'obiettivo.

Tuttavia, secondo Dennett, questo argomento può anche essere falso: è assurdo scartare l'ipotesi che un giorno l'IA riesca a realizzare un programma del genere, perché in linea teorica non è impossibile. Questa è la vera scommessa dell'IA.

I modelli finora approntati dagli studiosi di intelligenza artificiale mantengono per ora una loro valenza scientifica: essi sono utili nello studio dell'essenza della mente umana, e sono realizzati nell'ottica di una ultrasemplificazione, esattamente come i modelli utilizzati da altre branche della scienza<sup>14</sup>.

I poteri causali specificati da Dennett avrebbero quindi il merito di essere scientificamente comprensibili; ma questo dipende dal fatto che egli non sente la necessità, in quanto non lo reputa un problema reale, di dover postulare o spiegare l'esistenza di alcuni stati interni che, a suo avviso, si risolvono semplicemente nella loro componente oggettiva (o 'in terza persona'), ossia nel comportamento osservabile. Non c'è altro, secondo Dennett, da spiegare se non tali comportamenti.

Il punto cruciale per Searle, nella visione di Dennett, è in realtà il problema della coscienza: la coscienza introspettiva, ciò che si prova ad essere quel che si è, è il vero argomento di Searle.

All'origine del dibattito tra Searle e Dennett vi è proprio una totale divergenza sul significato che rivestono i concetti di coscienza ed intenzionalità. Nel paragrafo successivo, si tenterà di mostrare quali siano queste discordanze, con lo scopo di sottolineare come due differenti teorie dell'intenzionalità e della mente abbiano condotto a due approcci radicalmente differenti nel campo dell'IA.

## **I presupposti del confronto: coscienza ed intenzionalità tra Searle e Dennett**

Come si accennato alla fine del precedente paragrafo, Dennett ritiene che per Searle il vero nodo della questione non risieda nel problema dell'inderivabilità della semantica dalla sintassi, ma in quello che significa in prima persona essere coscienti di quello che si è.

In altre parole, "Searle ha evidentemente confuso una tesi concernente l'inderivabilità della *semantica* dalla sintassi con una tesi concernente l'inderivabilità della *coscienza della semantica* dalla sintassi. Per Searle, l'idea del genuino intelletto, della genuina 'semanticità' come spesso la chiama, non è scindibile dall'idea di coscienza"<sup>15</sup>.

Searle aveva definito la coscienza come un fenomeno biologico di ordine superiore causato dal cervello umano che, tra le sue varie caratteristiche strutturali, ne possedeva una, particolarmente importante, che fa sì che essa sia quello che è: la soggettività ontologica<sup>16</sup>.

---

<sup>13</sup> Dennett afferma che il cervello, che è dotato di milioni di canali, tutti capaci di attività simultanea, dimostra di possedere un'architettura massivamente parallela. Pertanto, è sua convinzione che i poteri causali necessari per controllare l'attività intenzionale esibita da esseri umani normali, possono essere ottenuti soltanto in un processore massivamente parallelo (ivi, p. 436).

<sup>14</sup> Questa posizione non è osteggiata da Searle; in questo senso, cioè come aiuto per studiare il funzionamento della mente umana, egli riconosce i meriti dell'IA. E' il ramo dell'IA definito "IA debole".

<sup>15</sup> D. C. Dennett, *L'atteggiamento intenzionale*, cit., p. 447.

<sup>16</sup> Ricordiamo che la "soggettività ontologica" è la caratteristica peculiare della coscienza, ciò che la rende più misteriosa: gli stati coscienti esistono solo in quanto esperiti da un agente, ossia possiedono quella che viene definita un'"ontologia in prima persona".

Dennett dissente radicalmente dalla concezione di coscienza così come la intende Searle; egli non pensa assolutamente che la coscienza contenga qualche elemento irriducibilmente soggettivo, perché questo la renderebbe inesplorabile da un punto di vista scientifico.

Ne "Il mistero della coscienza", e poi anche nella "New York Review of Books"<sup>17</sup>, si ritrova un acceso dibattito tra i due in merito al problema.

Il cuore della discussione è proprio l'elemento di soggettività che Searle individua come elemento fondante del fenomeno della coscienza.

A riprova e conferma del suo discorso, egli conduce i lettori a tentare alcuni semplici esperimenti su loro stessi, in modo che possano facilmente verificare cosa significhi 'essere coscienti'. Egli, ad esempio, suggerisce di provare a pizzicarsi la pelle; questo gesto provoca una sensazione di dolore, per quanto poco accentuata, che genera alcune conseguenze. Chiunque può verificare che tale sensazione di spiacevolezza possiede una particolare qualità soggettiva, che ciascuno percepisce in modo differente da ogni altro agente. Tuttavia, sottolinea Searle, la percezione di un dolore, come in questo caso, ha sia conseguenze epistemiche -ossia ognuno riconosce il proprio dolore come nessun altro può fare- ma la modalità con cui il dolore esiste è soggettivamente ontologica: esso esiste, cioè, solo in quanto esperito da un soggetto.

Questo significa, in generale, che la base neurofisiologica del cervello che causa gli stati coscienti è spiegabile da un punto di vista oggettivo o in terza persona; ma essi sono tali anche in ragione di un 'qualcosa in più' che si aggiunge loro, e questo qualcosa in più è proprio la qualità dell'esperienza soggettiva.

Insomma, per riprendere l'esempio precedente, secondo Searle, non si possono separare i processi neurofisiologici cerebrali che originano il dolore dalla qualità del dolore stesso, perché essi formano un tutt'uno che confluisce nella definizione di "dolore": "(...) i segnali che arrivano *causano* il dolore, e il dolore a sua volta provoca una disposizione comportamentale. Ma la cosa essenziale riguardo al dolore è che si tratta di una specifica sensazione interiore, qualitativa"<sup>18</sup>.

Ciò che Dennett osteggia di questa posizione è la sua non scientificità. Egli ritiene che la coscienza, intesa nel senso di Searle, non esista.

Non esiste per Dennett alcun problema dei 'qualia'<sup>19</sup> e della soggettività perché l'intera coscienza si risolve nel materialismo: in "Coscienza. Che cos'è"<sup>20</sup>, egli assimila la coscienza ad una macchina virtuale. L'immagine, che fa parte di una delle metafore tecnologiche di cui Dennett si serve, e che egli usa chiamare "pompe d'intuizione", vuole rimarcare proprio il fatto che non esiste un problema di irriducibilità della mente a qualcos'altro ('hard problem'), ma esistono solo delle difficoltà legate agli aspetti meccanici e tecnologici del funzionamento del cervello, ancora irrisolti ('soft problems')<sup>21</sup>.

L'obiettivo di Dennett è quello di comprendere e descrivere i fenomeni soggettivi di 'prima persona' dal punto di vista scientifico della 'terza persona'. Nella sua prospettiva, postulare delle qualità interne inaccessibili se non soggettivamente,

---

Searle aveva inoltre distinto due sensi in cui si può spiegare la distinzione oggettivo/soggettivo: il senso epistemico e quello ontologico. Il problema della coscienza è essenzialmente legato a quest'ultimo, che si riferisce al modo di esistere delle entità nel mondo.

<sup>17</sup> Una parte del dibattito è riportata in J. R. Searle, *Il mistero della coscienza*, tr. it. Raffaello Cortina, Milano 1998, pp. 93-107.

<sup>18</sup> Ivi, p. 78.

<sup>19</sup> Per la precisione, Dennett sostiene che l'equivoco si basi sull'uso di un termine fuorviante, 'qualia'. I 'qualia' non esistono, ma i fenomeni indicati dalla parola probabilmente sono reali, soltanto non vengono né compresi, né descritti con essa. Quindi, è opportuno ridefinire i 'qualia' in modo consistente e coerente. Dennett propone di paragonare i 'qualia' agli atomi e mostrare che anch'essi non sono semplici, ma complessi, e che sono composti di particelle non accessibili all'introspezione (cfr. E. Carli, *op. cit.*, p. 76).

<sup>20</sup> Cfr. D. C. Dennett, *Coscienza. Che cos'è*, cit., in E. Carli, *op. cit.*, p. 67.

<sup>21</sup> Con questa affermazione, Dennett ha come obiettivo polemico non solo Searle, ma anche quei funzionalisti che sostengono che vi sia qualcosa di irriducibile nella coscienza umana (→ vedi, ad esempio, D. J. Chalmers, *La mente cosciente*, tr. it. McGraw-Hill, Milano 1999).

significa ripristinare una sorta di dualismo cartesiano, proprio perchè tali stati interiori non sono indagabili oggettivamente e sperimentalmente.

Ed è precisamente questa l'accusa più forte che Dennett muove a Searle: quella, cioè, di avere alimentato e supportato proprio quella forma di dualismo<sup>22</sup> che egli asseriva di voler combattere.

Inoltre, Dennett non approva il metodo searliano: egli non si affida a veri metodi di ricerca, ma ad un'evidenza del singolo, ad una tradizione della psicologia del senso comune, che è tanto incontestabile quanto poco scientifica, in quanto non si basa su fenomeni osservabili e verificabili, passibili di essere generalizzati. Questo dissenso è originato dal fatto che, nella visione dennettiana, l'oggettività scientifica richiede il punto di vista in terza persona<sup>23</sup>.

Su quest'ultimo punto abbiamo già visto come si difende Searle, il quale suggerisce che la scientificità possa consistere anche di altri punti di vista; ossia, è per lui evidente che non esistono veramente solo i fenomeni verificabili scientificamente.

I fenomeni mentali, come abbiamo sottolineato prima, pur non essendo verificabili dalla prospettiva oggettiva, hanno comunque una modalità di esistenza inequivocabile, esistono cioè in relazione ad un soggetto che li esperisce. Ma esistono, ed è abbastanza semplice constatarlo: il problema della coscienza, perciò, per Searle, non si riduce solo alla comprensione degli aspetti meccanici del cervello, ma deve impegnarsi anche a spiegare le sensazioni soggettive, i 'qualia'.

Dennett, invece, secondo Searle, negando le sensazioni soggettive, nega i dati che una teoria della coscienza deve essere in grado di giustificare; egli offre una spiegazione alternativa di coscienza, secondo la quale l'essenza dell'esperienza cosciente risiede nelle relazioni causa-effetto tra alcuni input, o stimoli (ad esempio, nel caso del pizzicotto, la pressione sulla pelle), ed alcuni output, ossia disposizioni comportamentali con cui rispondiamo agli stimoli (ad esempio, il ritirare il braccio, o il massaggiarlo). Tra questi ultimi vi sono degli stati discriminativi che determinano a rispondere in modi diversi allo stesso stimolo, ma " il tipo di stato che abbiamo per discriminare la pressione è esattamente come lo stato di una macchina per determinare la pressione. Essa non prova alcuna particolare sensazione; in realtà non possiede alcuna sensazione interiore, perché le 'sensazioni interiori' non esistono"<sup>24</sup>.

Per Dennett, l'uomo stesso è una macchina; Searle pensa che la mente sia un calcolatore, ma solo a livello delle relazioni input-output, perché poi essa si distingue grazie alla capacità del cervello di causare stati mentali.

Ma per Dennett non c'è altro oltre ai poteri causali prodotti dal cervello che controllano e regolano l'attività intelligente esibita dal soggetto, e che si basano sulla giusta velocità di elaborazione delle informazioni e di gestione degli input e degli output. Dice Dennett: "Credo che alla gente non piaccia molto sentire dire che noi siamo soltanto delle macchine, ma questo è causato dal fatto che essi si riferiscono ad un'idea di macchina troppo semplicistica. (...) Sto dicendo che essi sono macchine estremamente complesse e sofisticate costituite di parti meccaniche e materiali, i nostri cervelli e i nostri corpi"<sup>25</sup>.

In "Brainstorms. Saggi filosofici sulla mente e la psicologia", Dennett mostra per quale ragione, nella sua prospettiva, un calcolatore non possa provare sensazioni, in particolare si rifà a a sua volta al classico esempio del dolore<sup>26</sup>. Poiché per Dennett la sofferenza non è il nome di una sensazione, ma significa vedere vanificare i propri piani e le proprie speranze, la questione di come un robot potrebbe sentire dolore non deve incentrarsi sul riferimento a qualcosa di 'soggettivo' e 'misterioso'. Il fatto che sino

---

<sup>22</sup> Il problema del dualismo tra mente e corpo ha origini molto antiche, ma viene consacrato dalla filosofia cartesiana. Cartesio, infatti, aveva separato la sostanza pensante ("res cogitans") dalla sostanza estesa ("res extensa").

<sup>23</sup> Questa posizione fa di Dennett, agli occhi di Searle, un verificazionista (cfr. J. R. Searle, *Il mistero della coscienza*, cit., p. 77).

<sup>24</sup> Ivi, p. 79.

<sup>25</sup> Intervista a D. C. Dennett, in E. Carli, *op. cit.*, p. 71.

<sup>26</sup> D. C. Dennett, *Brainstorms. Saggi filosofici sulla mente e la psicologia*, trad. it. Adelphi, Milano 1991, pp. 299-350.

ad oggi non si siano conseguiti dei successi in questa direzione, dipende solamente dal non possedere ancora una teoria coerente del dolore.

L'obiettivo di Dennett è quello di mostrare come la sofferenza, anche se non è direttamente osservabile, possa essere spiegata senza ricorrere alla postulazione di qualità interne o di proprietà intrinseche, come la 'terribilità', la disperazione o la tragicità<sup>27</sup>.

Dunque, per riassumere, secondo Searle gli aspetti computazionali di una macchina cosciente artificiale sarebbero qualcosa che si aggiunge alla coscienza, mentre per Dennett non esiste coscienza oltre alle caratteristiche computazionali, perché queste coincidono con la coscienza stessa.

Dennett taccia Searle di essere dualista, Searle accusa Dennett di antimentalismo. Quello che è interessante notare, è che una posizione come quella di Dennett non poteva che originare una versione dell'IA forte. Concepire la coscienza come corrispondente agli aspetti computazionali di una macchina cosciente artificiale non può che avvalorare la tesi per cui la mente sta al programma come il cervello sta all'hardware del calcolatore.

Ma le differenze radicali non sono finite qui. Al problema della coscienza si connette quello dell'intenzionalità, ed è quindi opportuno mostrare anche in questo caso quale sia il gap teorico che divide i due filosofi e che necessariamente li conduce distanti in merito alle posizioni dell'IA.

Abbiamo già accennato alla posizione di Dennett sull'intenzionalità, ma la riprendiamo ora alla luce della distinzione searliana tra intenzionalità intrinseca, derivata e come se. Secondo Dennett, un computer digitale elettronico dovrebbe attivare il programma giusto abbastanza velocemente da riprodurre in tempo reale il funzionamento del cervello e ciò che questo, tramite i suoi poteri causali, può produrre, ossia il controllo dell'attività intenzionale rapida e intelligente, esibita dagli esseri umani.

Abbiamo già sottolineato in che modo Dennett descriva i "poteri causali"; rimane ora da capire cosa egli intenda per "attività intenzionale rapida e intelligente".

Searle aveva distinto tre tipi di intenzionalità: intrinseca, derivata e 'come se'. Quest'ultimo tipo non è una vera e propria forma di intenzionalità; si tratta di un'attribuzione di intenzionalità che viene attuata verso un agente che si comporta, appunto, come se possedesse intenzionalità, senza di fatto averla.

Ciò che è rilevante in questa sede è la nozione di intenzionalità derivata.

Searle aveva sostenuto che alcuni nostri manufatti, così come le proposizioni del linguaggio, possono avere un'intenzionalità derivata<sup>28</sup> da noi, che, a nostra volta, possediamo un'intenzionalità intrinseca (o originaria), assolutamente non derivata.

Ne "L'atteggiamento intenzionale"<sup>29</sup>, così come ne "La mente e le menti"<sup>30</sup>, Dennett avanza la tesi per cui l'unico tipo di intenzionalità esistente è l'intenzionalità *derivata*.

Per avvalorare questa posizione, Dennett porta l'esempio di una macchina, piuttosto semplice: si tratta di un distributore automatico di bibite dotato di un particolare congegno (definito 'two-bitser') avente il compito di accettare e respingere quarti di dollaro americani. Esso deve essere costruito in modo tale da essere in grado di

---

<sup>27</sup> Dennett ha spesso portato l'esempio degli zombies a sostegno di questa tesi. Nella sua prospettiva, la sofferenza degli zombies non sarebbe differente dalla nostra sofferenza cosciente, essa non sarebbe meno importante della nostra. Secondo Searle, invece, ciò che Dennett esclude dalle sue indagini e che fa sì che l'uomo soffra e lo zombie no, è il fatto che l'uomo possiede un elemento di sensibilità che, per definizione, lo zombie non possiede (cfr. J. R. Searle, *Il mistero della coscienza*, cit., p. 86).

<sup>28</sup> Secondo Searle, l'intenzionalità intrinseca è indipendente dagli osservatori, sussiste di per sé in quanto elemento fondante dei fenomeni psichici: se si possiede uno stato mentale lo si ha indipendentemente da quello che gli altri pensano al riguardo. L'intenzionalità derivata, invece, è dipendente dagli osservatori: deriva da quella intrinseca, ed è frutto di una "imposizione" degli uomini ai loro manufatti e alle proposizioni del linguaggio. I manufatti e le frasi hanno intenzionalità derivata nel senso che significano qualcosa solo in relazione all'uomo che li ha creati.

<sup>29</sup> Cfr. D. C. Dennett, *L'atteggiamento intenzionale*, cit., cap. VIII.

<sup>30</sup> Id., *La mente e le menti. Verso una comprensione della coscienza*, tr. it. R. C. S. Libri & Grandi Opere, Milano 1997, pp. 62-68.

riconoscere un quarto di dollaro americano da qualsiasi altro pezzo tondo di metallo, indipendentemente dal suo valore.

Questo, naturalmente, finchè restiamo in America; ma, nel caso in cui si cambiasse Paese, lo stesso congegno dovrebbe essere strutturato in modo da accettare la moneta corretta rispetto al posto in cui ci si trova, e scartare le altre possibilità.

Ciò che Dennett vuole rimarcare è che la macchina non è dotata di una funzione intrinseca sua propria che non muta a seconda delle circostanze; al contrario, essa cambia in relazione alle intenzioni di chi la progetta e la realizza.

Si tratta dunque, evidentemente, di un caso di intenzionalità derivata: la macchina *significa* qualcosa, solo perchè siamo noi a dirle cosa deve significare.

Fino a questo punto, la posizione di Dennett non diverge da quella di Searle, e neanche da quella di altri funzionalisti, come Fodor e Dretske<sup>31</sup>; il vero momento di distacco si ha quando egli afferma di “applicare le stesse morali, le stesse regole pratiche di interpretazione al caso dell’uomo”<sup>32</sup>. In sostanza, per Dennett, noi siamo dei ‘two-bitser’, solo che molto più sofisticati ed elaborati.

La teoria cui Dennett fa riferimento per supportare tale tesi è la teoria dell’evoluzione delle specie.<sup>33</sup>

Anche in questo caso, egli parte da un esempio, una sorta di ‘esperimento mentale’<sup>34</sup>: l’idea è quella di immaginare di voler sperimentare la vita nel XXV secolo e che, per raggiungere tale fine, sia necessario collocare il nostro corpo in un meccanismo di ibernazione. A questo punto, diventa indispensabile progettare un supersistema atto a proteggere la capsula in cui si trova il nostro corpo e a fornirle l’energia necessaria per 400 anni.

Si suppone, perciò di costruire un robot su cui collocare la capsula e di istruirlo in modo da garantirne la sopravvivenza, per esempio, programmando nuovi corsi di azioni, scoprendo strategie di ricerca delle risorse che gli permettano di usufruirne per primo, imparando dai precedenti errori, studiando meccanismi di difesa, magari per mezzo di ‘alleanze’ con altri robot simili, ecc.

Naturalmente, questo tipo di macchina è molto più complesso del two-bitser dell’esempio precedente; tuttavia, anche se funzionasse in modo molto credibile in qualsiasi test di Turing<sup>35</sup> venisse sottoposto, esso avrebbe soltanto un’intenzionalità derivata, esattamente come qualsiasi manufatto.

Secondo Dennett, è possibile asserire, a questo punto, che “la nostra stessa intenzionalità è esattamente come quella del robot, dato che il racconto di fantascienza che ho narrato non è nuovo; è solo una variazione della visione che Dawkins<sup>36</sup> dà di noi (e di altre specie biologiche) come ‘macchine di sopravvivenza’ progettate per prolungare il futuro dei nostri geni egoisti”; “(...), la nostra intenzionalità è pertanto derivata dall’intenzionalità dei nostri geni ‘egoisti!’”<sup>37</sup>.

---

<sup>31</sup> Questi, pur abbracciando la dottrina funzionalista, non riescono a negare l’esistenza di un’intenzionalità intrinseca.

Cfr. F. Dretske, “Machines and the mental”, in *Proceedings and Addresses of the APA*, 1985, vol. 59, pp. 23-33.

Cfr. J. Fodor, *Psicosemantica*, tr. it. Il Mulino, Bologna 1990.

<sup>32</sup> D. C. Dennett, *L’atteggiamento intenzionale*, cit., p. 394.

<sup>33</sup> Per un ulteriore approfondimento, vedi D. C. Dennett, *L’idea pericolosa di Darwin*, tr. it. Bollati Boringhieri, Torino 1997.

<sup>34</sup> Cfr. D. C. Dennett, *L’atteggiamento intenzionale*, cit., pp. 396-401.

<sup>35</sup> Ricordiamo che il test di Turing è un sistema elaborato al fine di stabilire se una macchina è in grado di pensare o meno: una persona si trova davanti ad un terminale e con la tastiera scrive delle domande e riceve delle risposte. Dall’altro capo del terminale ci sono una macchina ed un operatore umano che forniscono alternativamente le risposte alle domande. Se la persona non è in grado di discernere quando sta parlando con la macchina e quando con l’operatore umano, allora la macchina è intelligente.

<sup>36</sup> Cfr. R. Dawkins, *Il gene egoista*, trad. it. Mondadori, Milano 1992, cit. in Dennett, *L’atteggiamento intenzionale*, cit., p. 400.

<sup>37</sup> Ivi, pp. 399-400.

Il problema, ora, per Dennett è capire come sia possibile che la nostra intenzionalità derivi da entità, i geni, che possiedono evidentemente un'intenzionalità 'come se'<sup>38</sup>. Inoltre, noi non siamo, al pari del robot, il frutto di un progetto di ingegneria consapevole e previdente, ma il risultato di un processo di progettazione di cui beneficiano i nostri geni e che è completamente privo di un progettista deliberato e intelligente.

Per risolvere il problema ed eliminare la necessità di un artefice intelligente, Dennett mette in gioco la teoria della selezione naturale.

Sarebbe piuttosto sciocco, secondo Dennett, supporre che i geni siano abili ingegneri; essi sono privi di qualsiasi intelligenza, non possono rappresentare o calcolare niente, usufruiscono solamente del processo di progettazione. Dunque, "chi e che cosa fa il progettista? Madre Natura, naturalmente, o più letteralmente, il lungo, lento processo di evoluzione per selezione naturale"<sup>39</sup>.

In un certo senso, si potrebbe dire che è Madre Natura a possedere un'intenzionalità intrinseca o originaria, da cui deriva la nostra intenzionalità derivata, in quanto siamo manufatti da lei progettati; tuttavia, Madre Natura non è un artefice intelligente, che ha pianificato ogni cosa per arrivare dove siamo, ma coincide con un lento processo di evoluzione in cui ha parte attiva la selezione: questa può 'scegliere' un progetto per una ragione o per un'altra, senza 'rappresentare' la scelta o le ragioni.

Inoltre, "è solo relativamente a tali 'scelte' del progetto o scopi 'firmati' dall'evoluzione - *raisons d'être* - che possiamo identificare comportamenti, azioni, percezioni, credenze, o qualsiasi altra categoria della psicologia del senso comune"<sup>40</sup>.

Quindi, il tipo di intenzionalità che contraddistingue il robot è esattamente come il nostro tipo di intenzionalità, cioè derivata, e non esiste nessuna intenzionalità intrinseca agli stati mentali.

E così come noi facciamo delle attribuzioni funzionali verso ogni nostro manufatto, che possono variare a seconda del contesto d'utilizzo della macchina, analogamente anche Madre Natura non ha attribuito delle funzioni univoche ed immutabili alle sue creazioni, ma esse si determinano di volta in volta nell'interazione con l'ambiente, al solo fine di garantire una miglior sopravvivenza della specie<sup>41</sup>.

Perciò, quando Dennett sostiene che l'IA debba tentare di riprodurre i poteri causali del cervello che originano il controllo della rapida ed intelligente attività intenzionale, parla di qualcosa di molto diverso da ciò che intende Searle. Alla luce di ciò che si è detto finora, infatti, è chiaro che anche per Dennett probabilmente la sintassi non sarebbe sufficiente per la semantica; ma il punto sollevato da Dennett è che in realtà non esiste nessuna semantica nel senso in cui la concepisce Searle: "(...) la sintassi *concretata, attivata* -il 'programma giusto' in una macchina opportunamente veloce- è sufficiente per l'intenzionalità derivata, e questo è l'unico tipo di semantica esistente"<sup>42</sup>.

E' evidente, in conclusione, che per Searle, che crede nell'esistenza di un livello coscienziale ed intenzionale soggettivo ed intrinseco agli stati mentali dell'individuo, le possibilità dell'IA di riprodurre una mente umana basandosi esclusivamente sull'analogia di questa con un programma di computer risultano molto basse; d'altra parte, la posizione di Dennett si sposa invece molto bene con una versione dell'IA forte, in quanto, eliminando il piano della coscienza e dei 'qualia', e considerando gli uomini alla stregua di macchine solo molto più sofisticate, non esclude che un giorno si riesca a creare davvero una macchina che pensa in modo umano.

---

<sup>38</sup> Dennett suggerisce infatti che è utile applicare l'atteggiamento intenzionale per descrivere i processi biologici; gli studiosi danno un senso a questi ultimi fornendoli di interpretazioni mentalistiche, ma l'intenzionalità che viene attribuita è una mera intenzionalità 'come se'.

<sup>39</sup> D. C. Dennett, *L'atteggiamento intenzionale*, cit., p. 401.

<sup>40</sup> Ivi, p. 402.

<sup>41</sup> Gli evoluzionisti hanno studiato proprio questa 'indeterminatezza della funzione'. Jay Gould, ad esempio, ha analizzato la funzione del pollice del panda: esso, anatomicamente, non è affatto un dito, ma un osso sesamoide del polso, spinto a servire 'come' pollice (cfr. S. J. Gould, *The panda's thumb*, W. W. Norton and Co., New York 1980, p. 22, tr. it. *Il pollice del panda*, Ed. Riuniti, Roma 1984).

<sup>42</sup> D. C. Dennett, *L'atteggiamento intenzionale*, cit., pp. 448-449.

## Origine e sviluppo dell'IA.

A questo punto della nostra trattazione, si profila la necessità -al fine di completare l'analisi condotta sinora- di presentare gli aspetti caratteristici dell'intelligenza artificiale, mostrandone l'iter evolutivo ed accennando ai risvolti più 'tecnici' delle scoperte e delle conquiste maggiormente rilevanti attuate in questo settore negli ultimi cinquant'anni.

I punti che verranno toccati riguardano la definizione e i fondamenti teorici dell'IA, un breve cenno alle tappe storiche più salienti e alle correnti di studio presenti all'interno della ricerca (ad esempio, le reti neurali e la vita artificiale).

Il capitolo si chiude con un paragrafo riguardante le nuove tendenze e le aspettative che si profilano, ossia quelle inerenti alla possibilità di sviluppare una coscienza artificiale.

## Gli obiettivi di fondo della ricerca sull'IA.

L'obiettivo di fondo della ricerca sull'Intelligenza Artificiale, in generale, è quello di costruire computers almeno in parte intelligenti.

Come abbiamo accennato, questo progetto ha incontrato diverse difficoltà: non si è rivelato semplice creare una macchina che riproducesse il comportamento intelligente, soprattutto quando si è cercato di ricreare al computer l'intelligenza comune, cioè la capacità che guida il comportamento quotidiano di ogni uomo.

In particolare, l'IA ha trovato dei problemi nella robotica, ossia nella costruzione di macchine intelligenti e capaci di interagire con il mondo esterno, nel riconoscimento del parlato e nella comprensione del linguaggio naturale.

Tuttavia, nonostante gli indubbi ostacoli pratici, i sostenitori dell'IA ritengono ancora che, per quanto complicato, non sia impossibile realizzare una macchina intelligente; posizione questa, che trova diversi oppositori tra cui, ricordiamo, John R. Searle, che ha tentato di confutare i fondamenti teorici stessi del progetto dell'intelligenza artificiale. Secondo l'IA, la mente, il pensiero, è una sorta di 'programma' fatto funzionare dal nostro cervello: come un calcolatore esegue il sistema operativo e i vari programmi applicativi, così il cervello umano "realizzerebbe" dei programmi per il linguaggio, per il ragionamento, ecc.

Come sappiamo, per i computers, la capacità di eseguire programmi non dipende dai materiali con cui essi vengono costruiti, ma soltanto dalla loro organizzazione logica.

I computers moderni si basano su piccolissimi microcircuiti stampati su una lastra di silicio, e si sono avviati degli studi per creare computers 'molecolari' o 'quantici'.

Tuttavia, la maggior parte degli studiosi del cervello è scettica sul fatto che il pensiero e la mente possano essere analizzati senza fare riferimento al modo in cui funziona questo organo: è infatti grazie ad esso ed alla sua elaborata struttura che gli esseri umani possono pensare.

Il cervello è costituito da un'enorme quantità di cellule, i neuroni, collegate tra loro in modo da costituire una rete tridimensionale lungo la quale si scambiano incessantemente segnali. Molti studiosi ritengono che una macchina, per essere intelligente, dovrebbe simulare questo processo mediante delle reti neurali artificiali.

Questo metodo di studio dei sistemi intelligenti è piuttosto recente e viene chiamato connessionismo; accanto ad esso, ha assunto una discreta importanza, nell'ultimo decennio, una disciplina definita Artificial Life, che, come vedremo, affronta il tema dello studio e della simulazione informatica dei sistemi biologici e dell'evoluzione.

## Definizioni e fondamenti teorici dell'IA.

In generale, con Intelligenza Artificiale si intende la progettazione di sistemi artificiali capaci di prestazioni paragonabili a quelle umane nello svolgimento di attività 'intelligenti' (tipicamente, la risoluzione di problemi e la comprensione del linguaggio naturale). Per la precisione, l'IA non è considerabile come una disciplina scientifica, ma

come un'area di ricerca tecnologica che nasce dall'incontro di varie discipline - psicologia, logica, linguistica- che ruotano intorno all'informatica.

Tale settore è diviso in due aree fondamentali: la prima è chiamata Intelligenza Artificiale Forte, e sostiene la possibilità di creare un computer che, correttamente programmato, possa essere dotato di un'intelligenza pura, non distinguibile in alcun modo rilevante da quella umana<sup>43</sup>.

La seconda area è definita Intelligenza Artificiale Debole e si contrappone nettamente alla prima, in quanto esclude la possibilità che un computer sia in grado, per quanto sofisticato esso sia, di eguagliare una mente umana; al massimo, potrà solo simulare alcuni processi cognitivi umani senza riuscire a riprodurli interamente.

Oltre a queste brevi definizioni, è opportuno anche mostrare quali siano i presupposti teorici dell'Intelligenza Artificiale; sono stati enucleati alcuni fondamenti, che riassumiamo qui di seguito, generalmente considerati alla base della moderna IA:

- il ragionamento, e in senso più ampio ogni attività della mente, è equivalente ad un calcolo;
- il calcolo può essere definito come manipolazione di simboli in base a regole prestabilite;
- il simbolo, o rappresentazione<sup>44</sup>, è un oggetto che sta per, o raffigura, un altro oggetto;
- può esistere un manipolatore automatico di simboli.

La nozione di calcolo, incentrata sul concetto di manipolazione di simboli in base a determinate regole, caratterizza il funzionamento di alcuni sistemi, denominati 'sistemi formali'<sup>45</sup>: essi lavorano su un insieme finito di elementi di base formalmente distinti utilizzando un insieme finito di regole che indicano le possibili trasformazioni di quei simboli. Molti giochi sono sistemi formali: ad esempio, la dama, gli scacchi e il backgammon; secondo i teorici dell'IA anche il pensiero umano può essere descritto come un sistema formale.

In questa direzione, uno dei primi obiettivi dell'IA è stato approntare una macchina che funzionasse come un sistema formale: le caratteristiche di una macchina simile sono state delineate dal matematico inglese Alan Turing (1912-1954).

## Cenni storici

L'IA nasce negli anni Cinquanta, come risultato della comprensione delle possibilità di elaborazione inerenti al calcolatore digitale di A. Turing e J. von Neumann. Era abbastanza chiaro fin da subito, infatti, che l'elaboratore non era di per sé confinato all'elaborazione di dati numerici, ma poteva elaborare dati di qualsiasi tipo, opportunamente codificati.

L'espressione 'Artificial Intelligence' è stata utilizzata per la prima volta in ambito scientifico da John McCarthy durante il convegno tenutosi a Darmouth nel 1956<sup>46</sup>.

In tale conferenza, si è mostrato che le macchine possono eseguire compiti la cui soluzione è considerata segno di intelligenza, come giocare a scacchi o dimostrare teoremi; nello stesso anno, a Boston, si è tenuta un'altra conferenza in cui diversi studiosi di differenti discipline erano concordi nel proporre un modello della mente basato sulla postulazione di processi interni che elaborano simboli o informazioni: è questo il momento di nascita della 'psicologia cognitiva', o 'cognitivismo', che si colloca alla base dei moderni studi sull'intelligenza artificiale.

---

<sup>43</sup> L'idea di base di questa teoria si ritrova nella filosofia di T. Hobbes, il quale riteneva che pensare equivalesse a calcolare: in quest'ottica, la mente umana sarebbe la risultante di un complesso insieme di calcoli eseguiti dal cervello (cfr. T. Hobbes, *Il Leviatano*, tr. it. G. Micheli, La Nuova Italia, Firenze 1976).

<sup>44</sup> Le rappresentazioni sono definite 'analogiche' se sussiste un rapporto di somiglianza con l'oggetto che raffigurano, oppure 'analitiche' se vengono realizzate mediante formule matematiche.

<sup>45</sup> La nozione di 'sistema formale', quale oggi la conosciamo, è dovuta agli studi di D. Hilbert e della sua scuola.

<sup>46</sup> Per la precisione, già nel 1952 M. Minsky aveva parlato di 'intelligenza artificiale' nella sua tesi presentata per il dottorato di Matematica presso l'Università di Princeton.

In particolare, la versione del cognitivismo incentrata sull'analogia mente-computer viene designata Human Information Processing (HIP). All'interno di tale paradigma, la mente è concepita come un'entità che filtra, seleziona, riorganizza e trasforma i dati che le provengono dall'esterno attraverso operazioni di tipo computazionale.

I primi modelli di funzionamento mentale suggeriti in questo ambito negli anni Sessanta sono contraddistinti da un'elaborazione di tipo 'seriale' delle informazioni<sup>47</sup> e dalla collocazione finale, nel processo di elaborazione, delle fasi di selezione. Il limite di tali modelli è che possiedono una capacità limitata di elaborazione dell'informazione, ma hanno il merito di essere molto semplici.

Tuttavia, non essendosi sempre rivelati validi, negli anni Settanta iniziano a nascere i modelli in parallelo<sup>48</sup>: in questi ultimi, l'elaborazione delle informazioni avviene contemporaneamente lungo canali in comunicazione tra di loro e le operazioni di selezione si trovano nelle prime fasi del processo. Essi possiedono una capacità illimitata di elaborazione, la possibilità di interazione tra i vari livelli dell'elaborazione dell'informazione e la possibilità di fare uso di strategie alternative.

Alcuni studiosi ritengono che l'architettura dei processi del cervello si basi essenzialmente sul parallelismo dei neuroni, tra cui è distribuita l'elaborazione dell'informazione. La mente strutturerebbe 'blocchi' d'informazione a partire dai dati percettivi immediati, la cui configurazione globale risulta in modo distribuito dall'attività parallela delle sue unità di elaborazione; il cervello funziona ad una molteplicità di livelli che lavorano anche contemporaneamente, per questo diversi teorici hanno pensato che potesse essere riprodotto da un calcolatore con un'architettura in parallelo.

Vi sono tre modelli di elaborazione parallela<sup>49</sup>: il modello vettoriale, il modello connessionista e il modello multiagente; le reti neurali costituiscono un esempio di tali modelli.

Esse sono state utilizzate da un'altra area di ricerca fondamentale dell'Intelligenza Artificiale, il 'connessionismo'.

Esso sostiene che per ottenere un comportamento intelligente da parte di una macchina occorre riprodurre o, almeno, simulare il funzionamento del cervello a livello cellulare; per fare ciò il connessionismo fa ricorso appunto alle reti neurali. Vedremo più avanti in cosa consiste esattamente questo settore di ricerca, ora ritorniamo all'inizio e vediamo quali sono stati i primi programmi creati e studiati dall'IA.

## **I primi programmi di IA**

A partire dalla metà degli anni Sessanta, grazie allo sviluppo della tecnologia dei circuiti integrati, nasce quella che viene chiamata la 'terza generazione dei computers'. Con tali nuove macchine, si è assistito ad un miglioramento dei programmi e all'indirizzarsi delle ricerche dell'IA soprattutto verso la comprensione del linguaggio umano, il calcolo simbolico e la rappresentazione semantica.

Nel 1963, Samuel ha scritto un programma capace di giocare a dama, dimostrando anche abilità nell'apprendimento: esso, infatti, utilizzava l'esperienza tratta durante le partite per perfezionare le proprie prestazioni successive.

Uno dei primi programmi di IA che ha avuto un discreto successo è *Eliza*, il cui nome fa riferimento a Eliza del "Pigmaliote" di G. B. Shaw, progettato da J. Weizenbaum nel 1964.

Tale programma era in grado di affrontare dei colloqui in inglese su 'copioni' prestabiliti di diverso argomento. Esso è noto, in particolare, per il ruolo di psicanalista di cui

---

<sup>47</sup> Questo significa che l'informazione, prima di accedere ad uno stadio successivo, deve essere stata completamente elaborata nello stadio precedente. Per le loro caratteristiche, questi modelli sono definiti modelli "ad oleodotto".

<sup>48</sup> Essi sono definiti anche modelli "a cascata".

<sup>49</sup> Cfr. O. Bettelli, *Modelli computazionali della mente*, Carabba, Lanciano 2002, p. 41.

Weizenbaum lo ha investito: *Eliza* lo svolgeva in modo così realistico da far pensare agli studiosi di essere di fronte ad un vero medico.

Sull'onda di questo esempio, K. Colby, uno psichiatra, ha inventato un altro programma simile, chiamato *Parry*, che riusciva a riprodurre il comportamento linguistico di un paranoico.

Successivamente, si è pensato di far interagire *Eliza* e *Parry*: ne sono risultati dei dialoghi giudicati dai medici molto realistici.

## Il linguaggio naturale in IA

Come abbiamo visto, uno dei limiti del computer e del suo modo di procedere è il fatto che esso non sa interpretare il significato dei simboli su cui lavora.

Già dagli anni Sessanta, si sono fatti dei tentativi per risolvere questo problema: tra il 1968 e 1970 T. Winograd<sup>50</sup> ha sviluppato il primo programma dotato di un modulo per la comprensione semantica del linguaggio umano, SHRLDU.

Esso era in grado di comprendere il linguaggio naturale e di spostare a piacimento, tramite un braccio articolato, un gruppo di oggetti geometrici differenti per forma, misura e colore. Precisamente, SHRLDU poteva esperire il mondo esterno attraverso le rappresentazioni proprie di oggetti fisici.

In seguito, sulla comprensione e la formulazione del linguaggio naturale sono stati fatti diversi studi, in particolare vi si è dedicato R. Schank a Yale alla fine degli anni Sessanta; sulla base di tali ricerche, nacque il programma MARGIE, proposto dallo stesso Schank nel 1975.

Il sistema si basava su modello di rappresentazione semantica chiamato dall'autore 'dipendenza concettuale'.

Sempre dalla scuola di Yale sono stati progettati altri programmi, come ELI, per perfezionare la fase di segmentazione del linguaggio naturale, e BABEL, per produrre testi.

I più recenti programmi sviluppati a Yale ('la terza generazione di computers'), come SAM, PAM, riescono a migliorare l'informazione aggiungendovi significati non esplicitati nel testo in esame.

Le ricerche sull'elaborazione del linguaggio naturale costituiscono uno dei settori di punta dell'IA e sono oggetto di un'ulteriore disciplina che si chiama linguistica computazionale e prendono avvio dalla teoria del linguaggio naturale di N. Chomsky<sup>51</sup>. Secondo Chomsky, la capacità linguistica che permette ad ogni essere umano di capire e formulare frasi nel suo idioma deriva da un insieme di conoscenze presenti nella sua mente, generalmente definite "competenza linguistica", in parte inconsapevoli.

Tale competenza si suddivide in tre sottoinsiemi, fonologica, sintattica, e semantica<sup>52</sup>; ciò che le accomuna è il fatto di nascere dalla presenza nella mente di un insieme di regole ben precise, analoghe alle regole proprie dei sistemi formali.

Uno dei formalismi più noti ed usati in IA per la rappresentazione delle conoscenze sono le *reti semantiche*, la cui prima formulazione è dovuta a Ross Quillian che l'ha enunciata nel 1968 per costruire un modello dell'organizzazione dei significati delle parole nella memoria e della capacità di associare concetti; ad esso sono seguiti vari tentativi di evolvere e migliorare il modello iniziale, pur non abbandonando le idee di Quillian.

Secondo queste intuizioni originarie, i concetti o significati delle parole nella mente sono interconnessi mediante una struttura reticolare, in cui esistono due tipi di 'nodi': i

---

<sup>50</sup> Cfr. T. Winograd, *Understanding natural language*, University Press-Academic Press, New York 1972.

<sup>51</sup> Cfr. N. Chomsky, *Reflections on language*, Pantheon, 1975.

<sup>52</sup> La competenza fonologica è inerente all'abilità umana di emettere e capire i suoni della lingua parlata; quella sintattica riguarda la capacità di formulare o distinguere una frase grammaticalmente corretta; infine, quella semantica concerne la capacità di assegnare o estrapolare il significato dalle frasi.

'nodi tipo', che corrispondono ai significati delle singole parole, e dai quali partono dei collegamenti o archi che finiscono nei 'nodi occorrenza'; questi descrivono il significato delle parole in esame tramite il loro stesso contenuto, ossia altre parole differenti.

Poiché anche ogni 'nodo occorrenza' è connesso al proprio 'nodo tipo' (quello che ne indica il significato), è evidente che nella rete ogni parola deriva il suo significato da un certo numero di altre parole.

Le reti semantiche di cui abbiamo parlato mostrano tuttavia due limiti: innanzitutto, ogni concetto è visto come un elemento unitario, in quanto l'articolazione del significato di un termine individuata dagli archi risulta troppo povera per spiegare tutte le componenti di un concetto. Inoltre, tale articolazione è considerata rigida, dal momento che ogni concetto viene descritto tramite una serie di componenti obbligatori (come succede nelle definizioni da dizionario).

La nostra memoria concettuale si dimostra estremamente più complessa rispetto a quanto le reti semantiche non possano spiegare; esse sono insufficienti per rappresentare la conoscenza.

Per tale ragione, M. Minsky ha elaborato un ulteriore modo per rendere conto del funzionamento della conoscenza e ha coniato la nozione di 'frame'<sup>53</sup>.

Un 'frame' è una struttura che raccoglie e dispone in base a vari livelli tutte le informazioni che concorrono a comporre un certo concetto; alcune di esse sono considerate necessarie, altre probabili e altre ancora solamente opzionali.

I 'frames', inoltre, sono connessi tra loro in modo tale che ogni elemento di ciascun 'frame' è collegato al 'frame' che ne definisce la struttura<sup>54</sup>.

Minsky ha anche ideato una teoria della Società delle Menti, in base alla quale la mente sarebbe composta da più 'agenti', organizzati gerarchicamente secondo regole precise, che comunicano tra loro per mezzo dei 'K-lines', sorta di canali di collegamento<sup>55</sup>.

## I sistemi esperti

I sistemi esperti sono considerati una delle più importanti applicazioni dell'IA.

Secondo una prima definizione semplificata, un sistema esperto è sostanzialmente un programma che è capace di dare una soluzione a problemi complessi rientranti in un ambito o dominio specifico; ad esempio, un sistema esperto può elaborare delle diagnosi mediche o valutare le possibilità di riuscita o meno di un'attività finanziaria.

All'origine del programma dei sistemi esperti, si pensa si trovino le idee alquanto originali ed innovative di K. Craik<sup>56</sup>, che hanno rappresentato una rivoluzione di pensiero fondamentale.

Secondo Craik, il cervello umano, che è un tipo particolare di macchina, costruisce modelli o rappresentazioni interne del mondo, e, tramite la loro elaborazione, determina delle azioni.

Queste, quindi, nell'automa ideato da Craik, non sono più la risposta meccanica a certi stimoli, come nell'automa cartesiano, in quanto è indispensabile il passaggio attraverso il livello intermedio della rappresentazione interna dello stimolo stesso e della sua elaborazione.

L'automa 'craikiano' possiede conoscenza e compie inferenze e l'intelligenza deve consistere nell'elaborare la conoscenza.

---

<sup>53</sup> Cfr. M. Minsky, *Semantic information processing*, the MIT Press, London 1968 e "Framework for Representing Knowledge", in Winston, 1975, tr. it. in J. Haugeland, *Progettare la mente. Filosofia, psicologia, intelligenza artificiale*, Il Mulino, Bologna 1989.

<sup>54</sup> Una nozione simile a quella di 'frame' è la nozione di 'script' formulata da Schank e Abelson; mentre nel 'frame' si rappresentano oggetti o stati di cose, lo 'script' permette di rappresentare eventi o sequenze di eventi tipici (cfr. R. C. Shank, R. P. Abelson, *Script, Plans, Goals and Understanding*, Lawrence Erlbaum, Hillsdale, (N. Y.) 1977).

<sup>55</sup> Cfr. Minsky, *La società della mente*, Adelphi, Milano 1989.

<sup>56</sup> Cfr. K. Craik, *The nature of explanation*, Princeton University Press, 1943, tr. it. parziale in V. Somenzi (a cura di), *La fisica della mente*, Boringhieri, Torino 1969, pp. 32-79.

Le idee di Craik, “spinte agli estremi, indicano infatti nella conoscenza la vera chiave dell’intelligenza: un sistema dotato di conoscenza in un certo dominio è in grado di risolvere i problemi che si verificano in quel dominio anche utilizzando mezzi logici molto limitati, mentre un sistema privo di conoscenza del dominio, pur se dotato di mezzi logici strepitosi, non è praticamente in grado di risolvere neppure i problemi più banali”<sup>57</sup>.

Per tali ragioni, la disciplina dei sistemi esperti viene definita come lo studio formale del comportamento umano in ogni dominio; essa è anche chiamata talvolta “ingegneria della conoscenza”, in quanto per realizzare un programma in un dato dominio è indispensabile possedere la “conoscenza del dominio”, cioè l’insieme di leggi che governano il comportamento di un agente in quel dominio.

Secondo McCarthy, la disciplina dei sistemi esperti deve soddisfare tre particolari requisiti essenziali:

- l’adeguatezza ontologica, cioè permettere la descrizione dei fatti salienti;
- l’adeguatezza epistemologica, cioè consentire l’espressione della conoscenza rilevante;
- l’adeguatezza euristica, cioè assicurare il compimento delle inferenze rilevanti.

Inoltre, egli ritiene che tali requisiti siano soddisfatti dal linguaggio della logica: la conoscenza rappresentata in questo modo costituisce un sistema formale di assiomi, partendo dai quali si possono dimostrare teoremi. Ogni insieme di conoscenza diviene perciò una teoria.

Dopo questa parentesi sulle caratteristiche e l’origine della disciplina dei sistemi esperti, torniamo ora alla definizione e proseguiamo con i componenti propri dei sistemi esperti e con alcuni esempi delle applicazioni più famose.

Generalmente, un sistema esperto deve possedere:

- a) una base di conoscenza specialistica su un determinato dominio, fondamentale per risolvere i problemi in quel settore; tale base deve essere adeguatamente rappresentata nella memoria del calcolatore;
- b) un motore inferenziale capace di dedurre, dalla base di conoscenza, le conclusioni, che consistono nelle soluzioni al problema facente parte del dominio; il motore inferenziale è il cuore del programma, e funziona applicando alla base di conoscenza delle procedure euristiche, simili a quelle utilizzate nel ‘problem solving’. Esse possono essere sia generali che specifiche per l’ambito di cui il sistema si sta occupando;
- c) un’interfaccia utente composta da alcuni moduli informatici che permettono all’uomo di interagire con il programma, ponendo domande e leggendo risposte.

Le prime applicazioni dell’IA ai sistemi esperti risalgono tendenzialmente agli anni Sessanta, in genere nelle università statunitensi.

Il termine “expertise”, ossia “esperienza scientifica”, è stato utilizzato per la prima volta nel 1967 da Joel Moses nella sua tesi di laurea, ma esistono dei precedenti nel programma MACSYMA del 1961, ancora basato su algoritmi tradizionali, e nel suo successore, DENDRAL, primo classico esempio di procedura euristica applicato al settore chimico.

Da quest’ultimo in avanti, i sistemi esperti sono stati impiegati essenzialmente nella risoluzione di problemi medici e, in tale ambito, va senz’altro ricordato MYCIN, creato nel 1972 all’Università di Stanford, che è stato il primo ad utilizzare la procedura di ricerca all’indietro nell’albero degli stati (‘backward chaining’).

## **Il connessionismo e le reti neurali**

In generale, come abbiamo visto, l’IA sostiene che l’intelligenza e le facoltà cognitive di un agente consistano nella sua capacità di manipolare automaticamente simboli.

Questo genere di sistemi, a partire dalla macchina di Turing fino ad arrivare al calcolatore digitale, non dipende assolutamente dal substrato materiale con cui viene realizzato: potrebbero essere usati, in modo equivalente, componenti elettronici al

---

<sup>57</sup> Cfr. P. Scaruffi, *La mente artificiale*, Franco Angeli, Milano 1991, cap. 3, “L’ingegneria della mente”.

silicio, componenti meccanici (leve, ruote, pulegge) o elettromeccanici; così come un computer potrebbe constare di molecole che interagiscono mediante reazioni chimiche, o da cellule organiche che si scambiano proteine ed enzimi.

Quello che è importante, in una macchina, è che possieda una struttura formale e funzionale.

Come sappiamo, l'IA trasferisce questi concetti allo studio della mente e suppone che i processi mentali non siano altro che il prodotto delle computazioni simboliche del cervello; anche in questo caso, il substrato materiale in cui tali processi avvengono sembra essere irrilevante ai fini della spiegazione dell'intelligenza.

Tuttavia, alcuni studiosi non sono concordi con questo approccio, e ritengono che sia al contrario indispensabile studiare la struttura biologica del cervello ed evidenziarne il funzionamento.

Recenti studi hanno dimostrato che il cervello sembra possedere un'architettura parallela, ossia i suoi meccanismi non sono equivalenti ad una macchina formale automatica a stati discreti e sequenziali, ma lavorano contemporaneamente: si tratta di miliardi di processi contemporanei caratterizzati da continue variazioni di stato.

Inoltre, alcune facoltà cognitive come la percezione non sembrano riducibili a processi computazionali simbolici.

Queste considerazioni, unitamente ai vani tentativi dell'IA di costruire modelli computazionali della mente effettivamente funzionanti, hanno preparato il terreno, alla fine degli anni Settanta, alla nascita di un nuovo orientamento nello studio dei processi mentali e nella costruzione di macchine intelligenti.

Questa nuova impostazione è definita *connessionismo*, e ha come fondamento teorico l'idea secondo la quale sia indispensabile emulare la fisiologia del cervello, cioè il funzionamento delle sue cellule, per spiegare e riprodurre l'intelligenza e le facoltà cognitive proprie dell'uomo o di altre specie animali.

Le ricerche nel settore connessionista si sono servite delle *reti neurali* per tentare di emulare il comportamento delle cellule neuronali.

Una rete neurale è una struttura formata da un certo numero di unità collegate tra loro da connessioni. Attraverso queste, ogni unità influenza fisicamente le altre con cui è connessa.

Le unità hanno alcune caratteristiche proprie delle cellule nervose, i neuroni del sistema nervoso reale, mentre le connessioni possiedono alcune delle caratteristiche essenziali dei collegamenti sinaptici tra neuroni<sup>58</sup>.

Una rete neurale è costituita da un insieme di nodi collegati; per ogni nodo, i collegamenti possono essere di due tipi, collegamenti di input - da cui arrivano segnali - e collegamenti di output- tramite i quali la rete emette segnali.

I nodi, inoltre, possono assumere due stati: lo stato di riposo e lo stato di attivazione, durante il quale il nodo invia segnali ai nodi cui è connesso.

Un aspetto particolare della fisiologia cerebrale emulato dalle reti neurali è il sistema di eccitazione: infatti, i collegamenti tra nodi possono essere o eccitatori o inibitori.

In linea generale, si può dire che il comportamento di una rete neurale consiste in un processo in cui, fornita alla rete come input una configurazione di segnali-stimolo (tramite l'attivazione dei nodi input), la rete produce come output un'altra configurazione di segnali.

Le reti neurali presentano anche alcune caratteristiche interessanti: in primo luogo, esse si differenziano a seconda della loro architettura, cioè in base alla disposizione (a strati) dei nodi; esistono perciò reti a uno strato, due strati, ecc.

In secondo luogo, le reti neurali si diversificano in relazione al tipo di collegamento tra i nodi: talvolta, le connessioni viaggiano solo in un senso; in altre situazioni, i messaggi vanno sia avanti che indietro; in altre ancora, nodi distanti possono essere collegati in modo diretto.

Una rete neurale può essere simulata 'via software', attraverso un normale calcolatore digitale, oppure attuata a livello hardware, attraverso la connessione di molti microprocessori che funzionano in parallelo.

---

<sup>58</sup> Cfr. D. Parisi, *Mente. I nuovi modelli della vita artificiale*, Il Mulino, Bologna 1999, p. 80.

Le reti neurali possiedono alcune proprietà che le rendono adatte a emulare il cervello umano e il comportamento intelligente; in particolare, se ne possono sottolineare tre:

-le reti neurali funzionano in modalità parallela, cioè in ogni momento molti nodi possono cambiare stato contemporaneamente. Come abbiamo visto, sembra che il cervello lavori in parallelo, perciò questo modello si dimostra più adeguato al funzionamento di quest'organo, rispetto al modello di elaborazione sequenziale proprio della macchina di Turing e von Neumann;

-le reti neurali sono in grado di apprendere a svolgere dei compiti, senza essere state esplicitamente programmate;

-nelle reti neurali, la conoscenza atta a eseguire un certo compito è diffusa in tutti i nodi della rete che la elaborano.

E' evidente, a questo punto, perché si dica che il connessionismo rappresenta una rottura nella tradizione classica dell'IA e della teoria computazionale della mente.

Senz'altro, lo ripetiamo, un motivo fondamentale è che il connessionismo sente la necessità e giudica prioritario emulare la fisiologia del cervello per produrre comportamenti intelligenti; ma il motivo primario è quello espresso dalla terza proprietà di cui si è parlato poco fa: l'intelligenza, i processi cognitivi, il pensiero, non derivano da processi computazionali simbolici, ma dal funzionamento globale delle unità subsimboliche che, prese da sole, non significano nulla.

Sulla base di questo principio, sono state realizzate alcune applicazioni delle reti neurali in alcuni ambiti, come nella simulazione dei comportamenti più semplici di specie inferiori, spesso affiancata dalla costruzione di robot reali. In questa direzione, il primo tentativo degno di nota è rappresentato dal 'Perceptron' di Rosenblatt del 1958, un robot capace di spostarsi in un ambiente eludendo gli ostacoli.

Le reti neurali si sono rivelate anche molto utili nello studio e nella emulazione dei processi che sottostanno alla percezione, settore in cui l'impostazione classica ha trovato parecchie difficoltà.

Importanti anche le applicazioni pratiche nel campo dei software per il riconoscimento dei caratteri e, recentemente, nel campo dell'analisi finanziaria.

Nonostante questi innegabili successi, il connessionismo presenta anche alcuni limiti, di difficile superamento: esso, infatti, è ancora lontano dall'essere applicato soddisfacentemente alla riproduzione artificiale delle facoltà cognitive superiori, questo anche in ragione del fatto che sono ancora scarse le conoscenze che abbiamo in proposito.

L'idea più diffusa è che tali facoltà, in qualche modo, consistano di qualche specie di elaborazione simbolica, non necessariamente computazionale, ma che ancora non si riesca a capire come questa scaturisca dall'interazione di molti processi di elaborazione paralleli e subsimbolici.

Tuttavia, per questi motivi, molti studiosi ritengono che il futuro dell'IA stia in una convergenza tra il modello computazionale e il modello connessionista: a grandi linee, si può affermare che il secondo si potrebbe applicare al 'livelli inferiori' dei processi cerebrali, in cui avvengono elaborazioni veloci e parallele di unità subsimboliche; e il primo si potrebbe occupare dei 'livelli alti', in cui si elaborano i simboli dopo averli ricavati dall'output dei livelli bassi.

## **La Vita Artificiale<sup>59</sup>**

Negli ultimi anni, oltre ai tentativi di riprodurre l'intelligenza e la mente tramite i calcolatori, alcuni ricercatori si sono dedicati ad una particolare disciplina, la *Vita Artificiale* (A. L., Artificial Life), che si può considerare la naturale estensione dell'intelligenza artificiale.

La Vita Artificiale, infatti, si pone come obiettivo la simulazione del comportamento di ecosistemi ed organismi reali tramite computer e, come l'IA cerca di capire i processi

---

<sup>59</sup> Il riferimento è a D. Parisi, *Mente. I nuovi modelli della Vita Artificiale*, cit., e a S. Silvi Antonini, *Vita artificiale. Dal Golem agli automi cellulari*, Apogeo, Milano 1995.

che sottostanno al funzionamento del cervello e della mente per riprodurre alcuni aspetti, allo stesso modo l'AL cerca delle risposte sulla vita degli organismi e sui principi evolutivi che hanno condotto alla formazione di esseri sempre più complessi servendosi di simulazioni create al computer.

Inoltre, la Vita Artificiale rivolge il proprio interesse allo sviluppo dei sistemi intelligenti, in quanto l'intelligenza è una caratteristica di alcune specie (in particolare quella umana) che si sono evolute a partire da organismi elementari: in questo senso, l'AL è concepita come un 'completamento' ed un approfondimento degli studi dell'IA.

L'AL è nata ufficialmente nel 1987 in una conferenza svoltasi all'Oppenheimer Study Center di Los Alamos (New Mexico), in occasione della quale il biologo C. Langton aveva riunito circa 160 studiosi appartenenti a diversi settori.

In pochi anni si sono raggiunti dei risultati interessanti, anche se allo stato attuale si è ancora fermi per lo più all'analisi di processi biologici elementari, come quelli propri dei virus o degli organismi monocellulari, o alla simulazione del comportamento di esseri che occupano un posto molto in basso nella scala evolutiva, come, ad esempio, gli insetti.

Generalmente, le creature studiate da AL sono assimilabili a simulazioni software che vivono in ambienti digitali, ma ci sono anche alcune applicazioni a piccoli robot, che sanno muoversi in ambienti reali, anche se semplici<sup>60</sup>.

Prima di descrivere due tra le più famose applicazioni di vita artificiale, ricordiamo che anche l'AL, come l'IA, è divisa in due posizioni: l'AL forte, la quale sostiene che le forme di vita artificiale e i loro modelli in natura sarebbero 'vivi' allo stesso modo; e l'AL debole, che asserisce solo la possibilità di comprendere meglio, tramite le simulazioni su computer, i misteri della vita, senza per questo pretendere che gli esseri artificiali possano definirsi viventi nel senso stretto del termine.

Automi cellulari→ sono i più noti enti artificiali studiati dalla vita artificiale. Un automa cellulare è una macchina software autodiretta che in genere 'vive' in ambienti simulati bidimensionali, esibisce comportamenti analoghi a semplici processi vitali e possiede la capacità di autoreplicarsi mettendo insieme elementi inerti presi dall'ambiente secondo regole basilari.

Inoltre, nel processo di replica, vengono simulate, tramite delle mutazioni, le variazioni degli organismi naturali teorizzate da C. Darwin. Così facendo, un sistema composto di automi cellulari simula il percorso evolutivo per selezione naturale, e la tempistica con cui questo si verifica è gestita e controllata dai ricercatori, differentemente da quanto accade negli esperimenti realizzati con esseri viventi.

Un esempio di applicazione dei principi degli automi cellulari è dato dalla creazione del matematico J. H. Conway, il quale ha ideato, nel 1960, un ambiente artificiale conosciuto con il nome di "Life" o "gioco della vita".

Tale ambiente è rappresentato da una scacchiera bidimensionale, in cui si trovano delle celle piene, che equivarrebbero agli organismi che vivono nell'ambiente; le celle vuote sono invece considerate morte. Tali organismi si evolvono in base a tre semplici regole, ciascuna delle quali simula processi che avvengono in natura:

- ogni cella che abbia adiacente un'altra sola cella o nessuna cella piene muore (simulazione della morte per isolamento);
- ogni cella che abbia almeno quattro celle adiacenti piene muore (simulazione della morte per sovraffollamento);
- ogni cella morta che abbia tre celle adiacenti piene torna in vita alla generazione successiva (regola di nascita).

Seguendo tali regole, il sistema si evolve autonomamente e realizza forme di vita complesse chiamate con strani nomi (lampeggiatori, alianti, bombardieri, fulmini); esse

---

<sup>60</sup> Un esempio è ancora il Perceptron di Rosenblatt, un robot tartaruga in grado di spostarsi evitando gli ostacoli (vedi paragrafo precedente).

poi mostrano comportamenti di vario genere, come dare la caccia alle forme più semplici, o difendere porzioni di territorio, ecc.<sup>61</sup>

algoritmi genetici → si tratta di un altro tipo di struttura evolutiva ed autoreplicantesi, ed è stata introdotta come idea per la prima volta da J. Holland a metà degli anni Settanta. Egli ha concepito tale struttura come un metodo per trovare ed ottimizzare gli algoritmi usati nel calcolo di alcune funzioni.

Per realizzare il suo scopo, Holland ha fatto riferimento al meccanismo della selezione naturale, applicato ad alcuni algoritmi definiti 'genitori', il cui patrimonio genetico 'digitale' è individuato da sequenze di bit.

A tali algoritmi vengono applicati due processi evolutivi: la mutazione casuale di uno o più bit del patrimonio genetico, e il crossing-over, cioè la generazione di un algoritmo 'figlio' che possiede un patrimonio genetico formato da una parte di quello dei due algoritmi 'genitori'.

Successivamente, si scelgono tra gli algoritmi 'figli' quelli più idonei alla risoluzione del problema affidato loro, e questi vengono sottoposti ad un nuovo processo di mutazione, mentre gli algoritmi 'peggiori' vengono eliminati.

Il processo va avanti finché non si trovano gli algoritmi in grado di risolvere il problema. E' interessante notare, in conclusione, che gli studi sugli algoritmi genetici vengono utilizzati anche in relazione a popolazioni di reti neurali che, come si è detto, possiedono capacità di autoapprendimento; in questo modo, sono state realizzate, ad esempio, delle 'formiche digitali' in grado di cercare cibo, lottare e abbattere i propri simili, evitare le sostanze velenose e pericolose.

Come si è visto, dunque, l'IA, e assieme a questa l'AL -intesa come suo completamento- ha tentato nel corso di questi anni di riprodurre l'intelligenza e il funzionamento della mente, facendo ricorso a diversi paradigmi di ricerca.

In generale, attualmente, si è piuttosto scettici riguardo alle possibilità di successo degli studi di IA, perché sembra che non si possa più prescindere, nello studio della mente e dei suoi meccanismi, dal tentativo di comprendere e replicare anche gli aspetti qualitativi dell'esperienza soggettiva, ossia quel complesso di percezioni definito 'coscienza'.

Insomma, l'uomo e la sua intelligenza non consisterebbero solo della giusta velocità di elaborazione delle informazioni e di processi cognitivi spiegabili come manipolazione meccanica e razionale di simboli, ma anche di una serie di elementi, apparentemente irrazionali, che però hanno una loro valenza nella costituzione dell'individuo, e che sembrano derivare dal funzionamento del cervello al pari di altri processi di pensiero.

Vediamo ora, in conclusione al presente lavoro, quali sono le tendenze verso cui si sta orientando l'IA e quali sono i suoi possibili sbocchi nel futuro, in relazione al problema - tornato in 'auge'- della coscienza e della necessità di riprodurre anch'essa artificialmente.

## **Dall'intelligenza artificiale alla coscienza artificiale**

Nel corso del presente lavoro, abbiamo visto quali siano i rapporti che intercorrono tra ciò che riguarda essenzialmente il mondo mentale-in particolare per quel che concerne intenzionalità e coscienza, concepite come caratteristiche intrinseche dei fenomeni psichici dell'essere umano- e il funzionalismo, soprattutto nel suo confluire nel campo di studi noto con il nome di 'Intelligenza Artificiale'.

Lo scopo dell'argomentazione consisteva nel mostrare quali fossero le principali tendenze presenti nel panorama della filosofia della mente e nel portare a soffermarsi sugli spunti di riflessione offerti dall'intero dibattito.

---

<sup>61</sup> Nel 1980, un altro matematico, S. Wolfram, si è interessato alla meccanica statistica degli automi cellulari lineari. Questi ultimi sono stringhe di celle estese all'infinito in entrambe le direzioni; in tali stringhe, in ogni istante, è attivo un numero finito di celle. Wolfram ha studiato e catalogato, numerandole progressivamente, le regole che ne governano il comportamento.

Il desiderio di arrivare a costruire artificialmente un'entità che al pari dell'uomo possa pensare, e anche esperire sentimenti, emozioni e vissuti, è diventato nel corso degli anni quasi un'esigenza, e da sogno fantascientifico quale era agli inizi- come insegnano i molti film e gli innumerevoli libri che si sono sviluppati intorno a questo tema- ha cominciato a rappresentare un obiettivo di ricerca su cui si lavora moltissimo attualmente, forse oggi anche più di un tempo, in seguito ad alcuni successi conseguiti che hanno portato un vento di ottimismo nel settore.

Tuttavia, fino a pochi anni fa, l'interesse generale, al di là di alcune eccezioni, era concentrato sui tentativi di studiare e riprodurre quel complesso di attività mentali che rendono alcune specie -e naturalmente, nella sua massima espressione, l'uomo- 'intelligenti'; indubbiamente, affrontare la coscienza da un punto di vista ingegneristico per molto tempo non ha interessato particolarmente la comunità scientifica internazionale.

Per queste ragioni, la coscienza, come si è spesso sottolineato, è stata relegata ai margini degli studi effettuati in IA, oppure, è stata ridefinita nella direzione di poter trascurare quegli aspetti qualitativi e soggettivi dell'esperienza che, per loro natura, si sottraevano ad una spiegazione oggettiva e scientifica.

Negli ultimi anni, però, la situazione si sta evolvendo verso orizzonti diversi: passata la prima ondata di entusiasmi per l'IA classica, si è notato che questa non era sufficiente per replicare un soggetto in tutti i suoi aspetti, e questo fallimento ha condotto sempre di più connessionisti e neuroscienziati verso la convinzione che fosse indispensabile conoscere esattamente i meccanismi di funzionamento del cervello per attuare una simile riproduzione.

Inoltre, si è gradualmente accettata l'idea che la coscienza, intesa non in senso morale o religioso, ma come capacità del cervello di determinare un soggetto unitario, in grado di fare esperienza di se stesso e del mondo circostante e produrre ragioni e scopi, valori soggettivi e sensazioni, non potesse più essere emarginata dai tentativi di comprendere l'essere umano, e soprattutto, fosse strettamente connessa al funzionamento cerebrale.

L'idea della 'multirealizzabilità' propria del funzionalismo, e poi dell'IA classica, secondo cui le attività intenzionali e mentali di un soggetto possono essere riprodotte, in quanto risultanti da una manipolazione formale di simboli, come se sussistessero da sole, ovvero indipendentemente dal substrato materiale su cui venivano sviluppate, ha perso diversi consensi, proprio in ragione del fatto che appare evidente che l'individuo è ciò che è grazie all'interazione di elementi tra loro inscindibili.

Gli aspetti qualitativi dell'esperienza (i 'qualia') esistono e derivano strettamente dall'attività del cervello, e la questione sembra essere molto più complessa di come veniva posta dall'intelligenza artificiale; finché non si conoscerà precisamente come funziona l'organo cerebrale, e quali meccanismi neurofisiologici sottostanno a determinati processi mentali (che possiamo tutti esperire nella nostra vita quotidiana), sono esigue le speranze di avvicinarsi all'obiettivo di creare un essere 'a nostra immagine e somiglianza'.

Le attuali tendenze sembrano perciò avvalorare le tesi di Searle, e riportare in 'auge' la coscienza e, con essa, l'intenzionalità, nella loro peculiarità ed unicità: in fondo, come la letteratura e la cinematografia fantascientifica hanno spesso mostrato, il vero desiderio dell'uomo sembra essere quello di arrivare a concepire un essere che abbia 'coscienza' più che 'intelligenza', in cui possa rispecchiarsi in quanto soggetto capace di provare dei sentimenti, inseguire obiettivi e prendere decisioni.

Poiché si tratta di un settore *in fieri*, è incerta la collocazione temporale della nascita di una 'artificial consciousness', sul modello della sua stretta parente, l'intelligenza artificiale (che, ricordiamo, prendeva avvio in una conferenza tenuta a Dartmouth nel 1956); tuttavia, un momento importante nella direzione di una riscoperta del valore della coscienza e dei suoi rapporti con l'intelligenza e con il cervello è rappresentato dal convegno di Tucson del 1994.

Questa data segna l'inizio di una serie di conferenze che, grazie all'intervento di neurobiologi e studiosi di scienze cognitive, si sono sempre più orientate verso una

maggior attenzione ed un approfondimento sul tema della coscienza, e, oggi, le conclusioni cui si è giunti non possono più essere facilmente ignorate.

Si ritiene anzi che il futuro dell'intelligenza artificiale si trovi nello sviluppo di una disciplina che, affinando una nuova teoria dell'intenzionalità e della coscienza, possa portare a replicare tutto quel complesso di vissuti che rendono tale l'essere umano.

Nel 1982, in Giappone, veniva inaugurato il programma di ricerca FGCS<sup>62</sup>, il cui obiettivo consisteva nello sviluppare le applicazioni di intelligenza artificiale; in seguito, nel 1987, al Convegno IJCAI di Milano l'oggetto del dibattito verteva proprio sugli interrogativi lasciati aperti dall'IA e sulle prospettive verso cui essa avrebbe dovuto rivolgersi.

Quello che emergeva con forza crescente era che le attività mentali non possono essere risolte esclusivamente in attività cognitive basate sull'elaborazione razionale e meccanica dei simboli; inoltre, veniva riconosciuta la stretta interrelazione tra coscienza ed intenzionalità: per questo all'inizio, gli sforzi erano protesi alla costruzione di un agente intenzionale<sup>63</sup>.

Insomma, la coscienza non è più un mero problema filosofico o terminologico, ma è diventata un problema scientifico: come sostiene Searle, gli aspetti qualitativi dell'esperienza cosciente esistono; sono comprovabili empiricamente, certo, ma questo non toglie loro valore o consistenza. Non si può spiegare la mente senza fare riferimento alla coscienza, e quindi è fondamentale superare il materialismo di cui la filosofia della mente è intrisa, e che è sorto come reazione estrema al dualismo di tipo cartesiano, per giungere ad un equilibrio diverso nel concepire i fenomeni mentali.

Questo è il prerequisito essenziale perché si possa pensare di realizzare un giorno un essere pensante e cosciente nel vero significato dei termini in questione.

Rimangono ancora aperti parecchi interrogativi, oltre a quelli più prettamente 'tecnici', legati alla attuali carenze conoscitive sul funzionamento del cervello o ai metodi con cui riprodurli.

I quesiti che si aprono riguardano le possibilità di successo di un programma del genere; anche supponendo di aver costruito un essere dotato di coscienza al pari dell'uomo, rimane il problema di determinare se effettivamente esso esperisce il mondo e se stesso come un essere umano. Le differenze sono molte anche all'interno dello stesso genere umano; alcuni processi sono universalizzabili, ma il modo soggettivo in cui il prodotto di tali meccanismi viene percepito può essere molto diverso: associamo, per esempio, alcune espressioni linguistiche a determinate sensazioni, e questo permette di comprenderci, ma a livello personale non è detto che ciò che significa 'essere innamorati' o 'provare dolore' per un agente sia *esattamente* la medesima cosa per un altro agente.

Diventa perciò difficile immaginare di costruire un essere che, pur manifestando comportamenti e parole indistinguibili da quelli umani, 'senta' come sente un essere umano; per lo meno, non è semplice verificarlo, forse anche perché, appunto, non esistono criteri universali o oggettivi con cui questo avviene, ma solo delle approssimazioni che permettono la comunicazione e la comprensione reciproca.

Bisognerebbe forse 'entrare' nelle altre menti, naturali o artificiali che siano, per coglierne certe sfumature e sapere davvero cosa provano di fronte ai medesimi stimoli; esse sono anche il frutto di uno sviluppo storico, sociale, educativo che impedisce una generalizzazione nel modo di percepire i vissuti e reagire alle situazioni presenti.

Detto questo, forse bisognerebbe anche chiedersi quale funzione ha l'intero organismo in tutto questo.

E' chiaro, e da millenni il cervello è stato rivestito delle funzioni di 'comando' dell'intero corpo umano (già in Galeno<sup>64</sup> si ritrovano barlumi di questa concezione), che l'organo cerebrale posseda un ruolo prioritario nel funzionamento del corpo e delle sue funzioni: soprattutto permette agli stimoli fisici del mondo esterno di essere in qualche modo elaborati, unificati, percepiti, ricordati, vissuti, ecc.

---

<sup>62</sup> Fifth-Generation-Computer-Systems.

<sup>63</sup> Si vedano le prime realizzazioni di G. Edelman.

<sup>64</sup> Galeno, Claudio (129-200 ca), medico e filosofo greco.

Tuttavia, nel momento in cui le attività cerebrali svolgono correttamente i propri compiti, rimane il fatto che il corpo ha, a sua volta, un ruolo non trascurabile: esso, infatti, veicola, attraverso i sensi, una serie di elementi che incidono sulle esperienze soggettive ed interiori.

Per esempio, il 'provare dolore' è un concetto ampio che comprende diversi livelli di manifestazione, ma la possibilità di 'sentire male' è determinata probabilmente anche dal fatto di avere una certa fisicità, di possedere una certa struttura organica, quindi di essere fatti di carne, sangue, terminazioni nervose, collocate in un certo modo preciso piuttosto che in un altro. E' da lì che parte il meccanismo; il corpo è il tramite di ogni esperienza, anche quella che sembra essere più soggettiva, come il provare un sentimento apparentemente sconnesso dagli stimoli fisici diretti.

Verrebbe quindi da domandarsi se sia sufficiente, nell'ottica di riprodurre un essere umano, replicare solo i meccanismi cerebrali o se, in una certa misura, non sarebbe necessario anche simulare la struttura ed il funzionamento biologico del resto del corpo.

Può essere che un ente costruito in altro modo possa 'provare dolore' se sottoposto a determinati stimoli, ma il fatto che sia composto da un materiale differente, seppur apparentemente analogo, porta a dubitare che possa trattarsi proprio di 'quel' dolore, o di qualcosa che possa definirsi dolore, nel senso in cui generalmente ne parliamo.

Forse sarà possibile, in futuro, realizzare un essere molto vicino all'uomo, ma è al momento difficile credere che si possa concretizzare un essere artificiale indistinguibile da quello umano; sono ancora troppe le domande, i dubbi, gli ostacoli che ci allontanano da quell'obiettivo.

Per ora, il sogno di una creatura artificiale dotata di coscienza, sentimenti, volizioni al pari di un uomo rimane quello che è stata finora: un sogno, forse un po' meno evanescente, ma pur sempre un sogno.

## BIBLIOGRAFIA

Bettelli, O., *Modelli computazionali della mente*, Carabba, Lanciano 2002.

Carli, E., *Cervelli che parlano*, Bruno Mondadori, Milano 1997.

Chalmers, D. J., *La mente cosciente*, tr. it. McGraw-Hill, Milano 1999.

Chomsky, N., *Reflections on language*, Pantheon, 1975.

Craik, K., *The nature of explanation*, Cambridge University Press, Cambridge 1943 (tr. it. parziale in V. Somenzi, *La fisica della mente*, Boringhieri, Torino 1969, pp. 32-79).

Dawkins, R., *Il gene egoista*, tr. it. Mondadori, Milano 1992.

Dennett, D. C., *Coscienza. Che cos'è*, tr. it., Rizzoli, Milano 1992.

Dennett, D. C., *L'atteggiamento intenzionale*, tr. it. Il Mulino, Bologna 1993.

Dennett, D. C., *L'idea pericolosa di Darwin*, tr. it. Bollati Boringhieri, Torino 1997.

Dennett, D. C., *La mente e le menti. Verso una comprensione della coscienza*, tr. it. R. C. S. Libri & Grandi Opere, Milano 1997.

Dennett, D., C., *Brainstorms. Saggi filosofici sulla mente e sulla psicologia*, tr. it. Adelphi, Milano 1991.

Di Francesco, M., *Introduzione alla filosofia della mente*, Carocci, Roma 2002.

Dretske, F., "Machines and the mental", in *Proceedings and Addresses of the APA* (1985), vol. 59, pp. 23-33.

Fodor, J., *Psicosemantica. Il problema del significato nella filosofia della mente*, tr. it., Il Mulino, Bologna 1990.

Gould, S. J., *The panda's thumb*, W. W. Norton and Co., New York 1980 (tr. it. *Il pollice del panda*, Ed. Riuniti, Roma 1984).

Gozzano, S., *Storia e teorie dell'intenzionalità*, Laterza, Bari 1997.

Hobbes, T., *Il Leviatano*, tr. it. di G. Micheli, La Nuova Italia, Firenze 1976.

Minski, M., "Framework for Representing Knowledge", in Winston, 1975 (tr. it. in Haugeland, *Progettare la mente, Filosofia, psicologia, intelligenza artificiale*, Il Mulino, Bologna 1989).

Minski, M., *La società della mente*, tr. it. Adelphi, Milano 1989.

Minski, M., *Semantic Information Processing*, The MIT Press, London 1968.

Parisi, D., *Mente. I nuovi modelli della vita artificiale*, Il Mulino, Bologna 1999.

Scaruffi, P., *La mente artificiale*, Franco Angeli, Milano 1991.

Searle, J. R., *Il mistero della coscienza*, tr. it. Raffaello Cortina, Milano 1998.

Searle, J. R., *La riscoperta della mente*, tr. it. Bollati Boringhieri, Torino 1994.

Searle, J. R., *Mente, cervello, intelligenza*, tr. it. Bompiani, Milano 1987.

Shank, R. C., Abelson, R. P., *Script, Plans, Goals and Understanding*, Lawrence Erlbaum, Hillsdale, (N. Y.) 1977.

Silvi Antonini, S., *Vita artificiale. Dal Golem agli automi cellulari*, Apogeo, Milano 1995.

Winograd, T., *Understanding natural language*, University Press-Academic Press, New York 1972.