Università Ca'Foscari Venezia

Corso di Laurea magistrale in Scienze Filosofiche

Tesi di Laurea

AI – Intelligenza Artificiale

"Ragionare non è nient'altro che calcolare"*, forse.

* Thomas Hobbes

filosofo e matematico inglese (Westport, 5 aprile 1588 – Hardwick Hall, 4 dicembre 1679)

Relatore	Ch. Prof. Enrico Jabara
Correlatore	Ch. Prof. Paolo Pagani
Laureando	Andrea Poser 803530
Sessione estiva, anno accademico	2019/2020

Indice

1 Introduzione 4
2 La meccanizzazione del pensiero 5
2.1 Aristotele 5
2.2 Leibniz 7
2.3 Boole 9
2.4 Frege 12
2.5 Gödel 14
2.6 Turing 22
3 L'intelligenza artificiale 28
3.1 Searle e l'argomento della stanza Cinese 32
3.2 La nazione cinese 37
3.3 Putnam e i cervelli nella vasca 37
3.4 AI forte e AI debole 40
3.6 Penrose 45
4 Libertà e determinismo 55
4.1 Spinoza 55
4.2 Spinoza e la logica di Boole 62
4.3 Chaitin 66
4.4 Successioni pseudo-casuali 68
5 Machine Learning 71
5.1 Gli algoritmi genetici 71
5.2 Reti neurali 75
6 Breve storia della sfida tra macchine e uomini 80
6.1 Deep Blue 80
6.2 Watson 81
6.3 Deep Mind 82
7 Conclusioni 83
8 Bibliografia 89

1 Introduzione

L'intelligenza artificiale (AI, l'acronimo inglese) è un campo di ricerca volto alla costruzione di animali artificiali (o almeno creature che, in certi contesti, sembrano essere animali) o persone artificiali (o almeno creature che, in certi contesti, sembrano essere persone).

La ricerca sulla possibilità dell'AI equivale, da un altro punto di vista, al chiedersi se l'intelligenza umana sia, almeno per principio, riducibile a una macchina per quanto evoluta e complessa questa possa essere.

Tali obiettivi sono evidentemente fonte di grande interesse e dibattito anche tra i filosofi nel tentativo di sostenere o escludere la replicabilità della mente.

Questo lavoro mira a presentare brevemente la storia della riflessione sul pensiero, sulle sue meccaniche interne (se di meccaniche si può parlare), i limiti teorici e la replicabilità, almeno teorica, dell'intelletto in una macchina. Infine verranno raccontate le ultime conquiste tecnico-scientifiche dei più avanzati computer.

Si precisa che questo lavoro non può ambire ad una trattazione approfondita e rigorosa di tutte le tematiche coinvolte ma spera comunque di poter mettere assieme qualche spunto di riflessione e approfondimento sui molti aspetti che hanno a che fare con l'intelligenza artificiale.

Il presente inizia dallo studio sulla logica del pensiero, o meglio sulla possibilità di tradurre in un sistema meccanico e rigoroso il pensiero. Questa storia parte della Grecia antica per poi evolversi con velocità sempre maggiore tra il XVII e il XX secolo.

2 La meccanizzazione del pensiero

2.1 Aristotele

Gli antichi greci, con Aristotele (Stagira, 384 o 383 a.C. – Calcide, 322 a.C.) pensavano che la parte più nobile dell'uomo, la parte che lo distingue dal resto degli altri animali, fosse la ragione. E per ragione si intende la capacità di eseguire dei ragionamenti corretti dal punto di vista logico. Aristotele ritenne di aver individuato dei principi logici universali che fossero a garanzia del corretto ragionamento. Come gli ingranaggi di una macchina, così i sillogismi sono quelle regole universali del corretto utilizzo della ragione.

Aristotele, negli *Analitici*, si sofferma sul procedimento di deduzione nel quale, da certe proposizioni date come premesse, ne segue una nuova e diversa proposizione.

I sillogismi rappresentano l'uso corretto del ragionamento e sono una connessione necessaria tra proposizioni e giudizi. La correttezza del ragionamento non dipende dalla verità delle proposizioni: infatti, anche connettendo correttamente proposizioni false possiamo giungere a giudizi logicamente corretti ma falsi. Si tratta quindi di uno schema formale che prescinde l'analisi delle premesse iniziali ma indica il modo corretto di utilizzarle. Tale attenzione alla forma, a prescindere dai contenuti delle proposizioni stesse, configura la logica aristotelica come la prima vera e propria logica formale nella storia del pensiero.

Il sillogismo si compone di due proposizioni che costituiscono le premesse ed una terza che ne è la conclusione. Le proposizioni possono essere affermative o negative, universali o particolari.

A loro volta le proposizioni contengono termini in funzione di soggetto o predicato.

Il classico esempio:

Premessa 1: *tutti gli animali sono mortali*

Premessa 2: *gli uomini sono animali*

Conclusione: *gli uomini sono mortali*

Il termine che ricorre nelle premesse ma non nella conclusione è detto termine medio (nell'esempio "animali") mentre gli altri termini presenti nelle premesse sono detti estremi. Più precisamente è estremo minore il termine che compare nella conclusione come soggetto ed estremo maggiore quello che compare nella conclusione come predicato. Ancora, la premessa che contiene l'estremo maggiore è detta premessa maggiore, mentre l'altra è detta premessa minore.

I sillogismi aristotelici si raggruppano in tre "figure" che si differenziano per la diversa posizione che il termine medio ricopre.

Nella prima figura del sillogismo il termine medio è soggetto nella premessa maggiore e predicato nella minore. I medievali la indicavano con l'abbreviazione *sub-prae*, da *subjectum-praedicatum*. L'esempio sopra riportato appartiene alla prima figura.

Nella seconda figura del sillogismo il termine medio compare due volte come predicato, sia nella premessa maggiore che nella minore. Per questa figura si usa l'abbreviazione *bis-prae*, ossia "due volte" (*bis*) "predicato" (*praedicatum*). Una delle premesse, in questo caso, è sempre negativa.

Premessa 1: *nessun asino è bipede*

Premessa 2: *ogni essere umano è bipede*

Conclusione: *nessun essere umano è un asino*

Nella terza figura del sillogismo il termine medio compare due volte come soggetto sia nella premessa maggiore che nella minore. È chiamato *bis-sub*, da "due volte" (*bis*) "soggetto" (*subjectum*).

Premessa 1: *qualche essere umano non è buono*

Premessa 2: *ogni essere umano è intelligente*

Conclusione: *qualcuno di intelligente non è buono*

Le possibilità della deduzione corretta offerta dalle tre figure si complicano e completano aggiungendo i modi, ovvero le varianti possibili a seconda delle forme affermative o negative di premesse e conclusioni e dell'universalità o particolarità delle stesse.

Non è qui il caso di esplicitare tutte le casistiche e le possibilità dei sillogismi; quel che è interessante sottolineare è che l'analisi svolta da Aristotele rappresenta il primo tentativo di ricondurre a regole logiche il ragionamento, nella convinzione che la ragione non sia altro che l'applicazione corretta di leggi logiche universali.

2.2 Leibniz

Nel 1642 Blaise Pascal aveva già presentato la *pascaline*, uno strumento di calcolo precursore della moderna calcolatrice che consente di addizionare e sottrarre numeri composti da un massimo di dodici cifre, operando automaticamente i riporti.

Nel 1673 il matematico Gottfried Wilhelm von Leibniz (Lipsia, 1 luglio 1646 – Hannover, 14 novembre 1716) progettò un nuovo modello di macchina calcolatrice capace di eseguire anche le moltiplicazioni e le divisioni. Ma l'idea di Leibniz va oltre l'automazione nel calcolo aritmetico spingendosi fino all'ipotesi di poter tradurre il ragionamento logico in uno speciale alfabeto. Come la sua macchina calcolatrice esegue meccanicamente le operazioni algebriche, così questo alfabeto consentirebbe di tradurre il pensiero umano in un sistema il cui il ragionamento diverrebbe un processo meccanico. Fondamentale è quindi la realizzazione, sul modello algebrico, di un sistema simbolico appositamente creato in cui ogni singolo simbolo rappresenta un'idea ben definita. Questo sistema, nell'ambizione del matematico, doveva abbracciare tutto il pensiero umano. Una sorta di estensione del modello algebrico è così presentata dallo stesso Leibniz: «sono sempre più convinto dell'utilità e realtà di questa scienza generale, e vedo che pochissime persone ne hanno compreso la portata … Questa caratteristica consiste in una certa scrittura o lingua … che rappresenta perfettamente le relazioni tra i nostri pensieri. I caratteri sarebbero diversissimi da tutto ciò che è stato immaginato finora; si è dimenticato, infatti, che i caratteri di questa nuova scrittura devono servire all'invenzione e al giudizio, come in algebra e in aritmetica. Tale scrittura avrà grandi vantaggi, ma fra gli altri ce n'è uno che a me sembra particolarmente importante, e cioè che usando questi caratteri sarà impossibile scrivere chimere come quelle che a volte ci si presentano alla mente. Un ignorante o non sarà in grado di usare questa scrittura, o cercando di usarla diventerà un erudito»[1].

L'idea di Leibniz, che egli chiamava *calcolus ratiocinator,* doveva essere preceduta dalla realizzazione di una sorta di enciclopedia della conoscenza umana nella quale ogni nozione potesse essere condensata in un simbolo adeguato. L'aggiunta infine di regole deduttive e di manipolazione di questi simboli avrebbe consentito la formalizzazione dell'intero pensiero umano e quindi la possibilità della sua trascrizione in una macchina, logicamente paragonabile a quella da lui inventata per somme, sottrazioni, moltiplicazioni e divisioni.

[1] Lettera di Leibniz a Jean Galloys del dicembre 1678 in Leibniz, G.W., *Mathematische Schriften*, a cura di C.I. Gerhardt, 7 voll., Berlin (Voll. I-II) e Hll (voll. III-VII), 1849-1863; rist. Anast. Georg Olms, Hildesheim, 1962.

2.3 Boole

Due secoli dopo Leibniz, il matematico inglese George Boole (Lincoln, 2 novembre 1815 – Ballintemple, 8 dicembre 1864) aggiunse un importante tassello nella storia della logica. Al concetto del sillogismo aristotelico aggiunse l'intuizione che i significati di parole come "animale" o "mortale" (come negli esempi sopra riportati) potevano esser definiti come *classe* o *collezione* di tutti gli individui descritti dalla parola stessa (la classe degli animali, la classe dei mortali) e tradusse il ragionamento in una nuova algebra: «… se come termine di una descrizione usiamo un aggettivo, per esempio "buono", con una lettera y rappresenteremo quelle cose alle quali può applicarsi la descrizione "buono", vale a dire "tutte le cose buone", o la classe "cose buone". Conveniamo inoltre di rappresentare con la combinazione xy la classe di cose a cui sono applicabili, simultaneamente, i nomi o le descrizioni rappresentati da x e y. Così se x da solo sta per "cose bianche", e y sta per "animali", xy starà per "animali bianchi"; analogamente, se z sta per "cose dotate di corna" … xyz rappresenterà "animali bianchi dotati di corna"»[2].

Nell'algebra di Boole:

$xx = x$	vale a dire l'unione di una classe con se stessa corrisponde alla stessa classe;
0	rappresenta una classe alla quale non appartiene nulla, classe vuota;
1	rappresenta tutti gli elementi che si considerano;
$x + y$	insieme delle cose presenti nella classe x o y. Rappresenta l'unione delle due classi;
$x - y$	tutte le cose presenti in x ma non in y.

---

[2] Boole, G., *An investigation of the Laws of Thought on whitch are fouded the mathematical Theories of logic and probabilities*, Walton & Maberly, London 1854; rist Dover New York, 1958 [trad. It. *Indagine sulle leggi del pensiero su cui sono fondate le teorie matematiche della logica e della probabilità*, Einaudi, Torino, 1976].

Quanto definito ci consente di scrivere la seguente equazione:

$$x + (1 - x) = 1$$

Inoltre, utilizzando l'algebra tradizionale in cui $xx = x^2$ possiamo trascrivere $xx = x$ in $x^2 = x$ oppure $x - x^2 = 0$ e ancora $x(1 - x) = 0$. Quest'ultima può esser letta come: niente può appartenere e non appartenere alla stessa classe. Questo risultato è la traduzione algebrica del principio logico di non contraddizione aristotelica.

Non solo il fondamentale principio di non contradizione può essere tradotto nella nuova algebra, ma anche i sillogismi. A seguire un esempio:

tutti i cavalli (x) *sono mammiferi* (y)

tutti i mammiferi (y) *sono vertebrati* (z)

tutti i cavalli (x) *sono vertebrati* (z)

Dire che tutto ciò che è in x è anche in y equivale a dire che non esiste nulla che appartenga a x ma non a y, ovvero $x(1 - y) = 0$ che diviene $x = xy$. Allo stesso modo per l'altro termine del sillogismo possiamo scrivere $y = yz$.

Il tutto può essere riassunto con:

$$x = xy = x(yz) = (xy)z = xz$$

I sillogismi rappresentano solo una parte della logica e Boole si rese conto che il proprio sistema andava oltre la traduzione matematica della logica aristotelica ed era utilizzabile anche a proposizioni secondarie, ovvero a proposizioni esprimenti relazioni fra proposizioni. Ne è un esempio la traduzione algebrica della prova dell'esistenza di Dio di Samuel Clarke[3]:

a) qualcosa esiste
b) se esiste, allora, o qualcosa è sempre esistito oppure le cose che esistono ora sono sorte dal nulla
c) se qualcosa esiste, allora, o esiste per necessità della propria natura, oppure per volontà di un altro essere
d) se esiste per necessità della propria natura, allora qualcosa è sempre esistito
e) se esiste per volontà di un altro essere, allora l'ipotesi che le cose che esistono siano sorte dal nulla è falsa.

Quanto sopra in simboli diviene:

$x = qualcosa\ esiste$

$y = qualcosa$ è $sempre\ esistito$

$z = le\ cose\ che\ esistono\ ora\ sono\ sorte\ dal\ nulla$

$P = il\ qualcosa\ di\ cui\ si$ è $parlato\ sopra\ esiste\ per\ necessit$à

$della\ propria\ natura$

$q = questo\ qualcosa\ esiste\ per\ volont$à $di\ un\ altro\ essere$

Ne seguono le seguenti equazioni[4]:

$1 - x = 0$ $\qquad$ $(x$ è $vera,\ x = 1)$

$x\{yx + (1 - y)(1 - z)\} = 0$

$x\{pq + (1 - p)(1 - q)\} = 0$

---

[3] Filosofo inglese (Norwich, 11 ottobre 1675 – Londra, 17 maggio 1729). Clarke Sosteneva che la matematica fosse sufficiente a dimostrare l'esistenza di Dio e delle leggi della natura. Ma la ragione non può giungere alla vera conoscenza di Dio ed è quindi un preambolo alla fede e alla rivelazione.

[4] In questo caso: l'equazione $x = 1$ indica che xè vera; $x = 0$ indica che xè falsa; $xy = 1$, quindi vera, se entrambe le variabili sono vere; $xy = 0$ se almeno una della variabili è falsa; $X = 0$ indica non x; $x(1 - y) = 0$ indica se $x$allora y (infatti se sostituiamo a $x$il valore 1 l'equazione diviene $1(1 - y) = 0$ ovvero $y = 1$.

$p(1-y)=0$
$qz=0$

Una complessa dimostrazione filosofica è stata tradotta in un insieme di formule relativamente semplice.

2.4 Frege

Nel 1879 Gottlob Frege (Wismar, 8 novembre 1848 – Bad Kleinen, 26 luglio 1925) pubblica *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens,* "Ideografia", linguaggio in formule del pensiero puro modellato su quello dell'aritmetica. Come già il titolo anticipa, l'obiettivo dell'autore era lo studio di un sistema logico che comprendesse tutte le inferenze deduttive utilizzate in matematica. Il logico tedesco riteneva fondamentale introdurre nuovi appositi simboli (rispetto a quanto fatto da Boole) affinché questa nuova sistematizzazione della logica potesse essere a fondamento sia dell'algebra che degli altri rami della matematica.
Frege introdusse i quantificatori universale ∀ ed esistenziale ∃.
Infatti nelle le frasi

Tutti i cavalli sono mammiferi – se x è un cavallo allora x è un mammifero

o

alcuni cavalli sono purosangue – x è un cavallo e x è un purosangue

la variabile x ha due significati distinti. Nel primo caso intendiamo che ciò che vogliamo dire è valido per tutti gli x, per tutti i cavalli; nel secondo caso solo alcuni x sono purosangue.
I quantificatori ci consentono una precisione logica importante trascrivendo le frasi sopra riportate in:

$$(\forall x)(se\ x\ \text{è}\ un\ cavallo, allora\ x\ \text{è}\ un\ purosangue)$$
$$(\exists x)(x\ \text{è}\ un\ cavallo\ e\ x\ \text{è}\ un\ purosangue)$$

Il sistema di Frege introdusse altri simboli quali

⊃	se… allora
¬	non…
∧	… e …
∨	… o …

che permise di trattare proposizioni non trascrivibili con la logica di Boole come ad esempio "*tutti gli studenti bocciati sono stupidi o pigri*":

$$B(x) per\ x\ \text{è}\ uno\ studente\ bocciato$$
$$S(x) per\ x\ \text{è}\ stupido$$
$$P(x) per\ x\ \text{è}\ pigro$$

Con il suo sistema Frege avrebbe quindi potuto tradurre la proposizione di esempio con:

$$(\forall x)(B(x) \supset S(x) \vee P(x))$$

Il logico tedesco di fatto costituì il primo esempio di un linguaggio formale artificiale. Una volta definite le *regole d'inferenza* - che riguardavano le configurazioni dei simboli - è possibile utilizzare operazioni tecniche per elaborare i simboli e di conseguenza i loro significati, indipendentemente dai significati stessi.

Ad esempio dati due qualsiasi enunciati α e β, se è vero sia β che (β ⊃ α) allora è vero anche α. La verità della conclusione è garantita solo dalla verità delle premesse e del corretto utilizzo della regola d'inferenza, indipendentemente dal significato che i simboli possono assumere.

Un problema fondamentale incrinò il sistema progettato dal logico tedesco. Egli infatti aveva basato anche i numeri naturali in termini puramente logici. Per far ciò assunse che un numero naturale è una proprietà di un insieme ovvero il numero dei suoi elementi. Ad esempio il numero 3 è per Frege l'insieme di tutti gli insiemi che posseggono la proprietà della "triplicità" ovvero che son composti da tre membri. Possiamo quindi definire in generale i numeri come totalità di insiemi equivalenti, ovvero di insiemi che possono essere messi in corrispondenza biunivoca tra loro (che significa che ad ogni elemento di un insieme corrisponde uno ed un solo elemento dell'atro insieme e viceversa). L'aritmetica di Frege si basava quindi su questi insiemi e insiemi di insiemi. Tale architettura è però minata alle fondamenta dal paradosso di Russell [5] che ne evidenzia la contraddittorietà.

2.5 Gödel

Kurt Friedrich Gödel (Brno, 28 aprile 1906 – Princeton, 14 gennaio 1978) è considerato uno dei più grandi logici di tutti i tempi.

Le difficoltà evidenziate da Russell si inserivano in un generale problema matematico circa la dimostrabilità interna (completezza e coerenza) di un sistema assiomatico. Questo infatti rappresentava uno dei 23 problemi proposti da David Hilbert (Königsberg, 23 gennaio 1862 – Gottinga, 14 febbraio 1943) nel secondo congresso internazionale di matematica tenutosi a Parigi nell'agosto del 1900.

Primo passo per questo ambizioso progetto è la formalizzazione completa del sistema in cui tutti i significati divengono irrilevanti per lasciare spazio solo ai segni e alle loro relazioni. Questi possono essere poi combinati con delle regole definite in modo tale che tutto sia perfettamente codificato. I postulati e teoremi di un sistema sono quindi sequenze di segni combinati secondo le regole stabilite. Tali catene e sequenze (di segni) sono strutture logiche prive di significato. Ciò nonostante è possibile fare delle affermazioni circa queste

[5] Ad esempio: definiamo classe un insieme o aggregato di oggetti distinti (ognuno dei quali è definito elemento della classe); una classe è detta normale se e solo se non contiene se stessa come elemento. Ora, se 'N' è la classe di tutte le classi normali, è 'N' una classe normale? La contraddizione che ne emerge evidenzia come non ci si possa fidare dell'apparente chiarezza di alcuni modelli per concludere sulla loro coerenza.

sequenze di segni. Tali affermazioni sono esterne al sistema formale stesso e rappresentano la metamatematica, ovvero considerazioni dal "di fuori" sul sistema formale (matematica senza significato). La distinzione è importante in quanto il progetto di Hilbert mirava a dimostrare la coerenza di un sistema di segni, una volta che questo fosse stato completamente formalizzato. Tale dimostrazione doveva basarsi su un numero finito di proprietà strutturali delle formule e su un numero finito di operazioni sulle stesse. Questi procedimenti vengono chiamati appunto "finitistici" e la relativa dimostrazione "assoluta". Per riassumere, l'obiettivo di Hilbert era di dimostrare la coerenza "assoluta" della teoria dei numeri una volta dato un insieme di regole di combinazione e di assiomi di partenza.

Le regole di combinazione dei segni possono essere le regole di inferenza logica. A riguardo si sottolinea come dai tempi di Aristotele fino alla metà dell'Ottocento non fossero stati compiuti passi avanti sull'analisi della logica e come le regole e i ragionamenti utilizzati comunemente dai matematici fossero di fatto regole non scritte su come ragionale. Solo nel XIX secolo con Boole si assistette a un'importante ripresa degli studi volti ad esplicitare i meccanismi del pensiero che normalmente utilizziamo "inconsapevolmente". In questo contesto si innestano i *Principia Mathematica* (A. Whitehead, B. Russell, *Principia Mathematica,* Cambridge University Press 1910-13), ovvero il tentativo di una soluzione finale ai problemi della coerenza dei sistemi matematici attraverso la riduzione degli stessi ai problemi di coerenza formale della logica, come se di fatto la matematica fosse un'applicazione della logica.

Si fornisce un esempio di dimostrazione assoluta di coerenza. Per far ciò è necessario:

- preparare un catalogo completo di segni, una sorta di vocabolario, che useremo nel calcolo;
- decidere le regole di combinazioni possibili dei segni che saranno quindi le "formule/proposizioni";

- definire le regole di trasformazione ovvero come trasformare formule in altre formule;
- scegliere alcune formule come assiomi e fondamenta del sistema stesso.

Da un sistema così descritto si può dimostrare che:

'$p \supset (\neg p \supset q)$' (se p, allora se non p allora q)

In base alla regola di sostituzione (che consente di sostituire una o più variabili con un'altra variabile) ed alla regola di separazione (che permette di dedurre una proposizione $S2$ da altre due proposizioni, una delle quali è $S1$ e l'altra $S1 \supset S2$), e supponendo che una certa formula S e la sua contraddizione $\neg S$ (sostituendole a 'p') siano deducibili dagli assiomi, ne consegue che sarebbe deducibile la formula 'q'. Il che equivale a dire che (sostituendo alla variabile 'q' una qualsiasi formula) qualsiasi formula è deducibile da un insieme incoerente di assiomi. Al contrario la presenza di una formula non deducibile dagli assiomi è la conferma della coerenza del sistema. Per dimostrare questa coerenza dobbiamo allora dimostrare che esiste una formula che non può essere dedotta dagli assiomi (coerenti). Per procedere a risolvere quest'ultimo punto è necessario individuare una caratteristica o proprietà delle formule che soddisfi le seguenti caratteristiche:

- la proprietà deve essere comune a tutti gli assiomi;
- la proprietà deve essere ereditaria ovvero trasferirsi dai teoremi alle formule correttamente dedotte dagli assiomi stessi;
- dobbiamo trovare una formula che non possiede la qualità in questione.

Stiamo insomma cercando una formula che sia sintatticamente corretta e che non possegga la proprietà, ereditaria, degli assiomi (e quindi da questi derivata). Come detto se gli assiomi fossero incoerenti ogni formula sarebbe un teorema e il trovare una formula che non sia un teorema e quindi una "conseguenza" dei teoremi rappresenterebbe la dimostrazione della coerenza degli assiomi stessi. La proprietà individuata a tale scopo è la tautologia (che in logica rappresenta una proposizione necessariamente vera – o piove o non piove). Si dimostra che gli assiomi sono delle tautologie e che questa caratteristica si trasmette ai teoremi da essi derivati. La formula $'p \vee q'$ è una formula, non è una tautologia e non è un teorema (ad esempio sostituendo le variabili: Carlo è alto o piove). Esiste quindi una formula che non è un teorema in quanto da questi non può essere derivata. Poiché ogni formula sarebbe un teorema se gli assiomi fossero incoerenti, il sistema è coerente e questa rappresenta una dimostrazione assoluta della coerenza del sistema. È utile aggiungere che è anche dimostrabile che tutte le tautologie sono dei teoremi ovvero che il sistema è *completo*, cioè che gli assiomi sono sufficienti per derivare tutte le tautologie (tutte le verità logiche del sistema). La prova di coerenza sopra riportata è relativa al sistema del calcolo proposizionale che rappresenta solo una piccola parte dell'aritmetica. L'obiettivo di Hilbert di estendere queste dimostrazioni a tutta l'aritmetica è destinato a fallire, come dimostrato da Gödel.

Nel 1931, Gödel dimostrò due cose:

- che è impossibile dare una dimostrazione della coerenza di un qualsiasi sistema contenente l'aritmetica;
- che un sistema contenente l'aritmetica è incompleto, ovvero esistono verità che non possono essere dimostrate a partire dai teoremi.

Si tratta, quindi, di certificare i limiti intrinseci ad un sistema assiomatico formale il quale non può essere in grado di dimostrare tutte le verità del sistema stesso e deve arrendersi quindi all'accettazione di verità indimostrabili (e tale conclusione non cambia anche se si ipotizza di aggiungere nuovi assiomi per tentare di colmare questa lacuna).

Gödel giunse a questi risultati con l'utilizzo di una speciale rappresentazione. Con il termine rappresentazione si intende un metodo attraverso il quale una struttura astratta di relazioni in un dominio è valida anche su di un altro dominio che però ha il vantaggio di semplificare e rendere più intellegibili le relazioni interne (valide in entrambi i domini). Esempi di rappresentazioni sono l'utilizzo dell'algebra per dimostrazioni geometriche, le realizzazioni di modelli in scala per anticipare i comportamenti degli oggetti con dimensioni reali, la geometria proiettiva che traduce ed evidenzia relazioni matematiche.

Il metodo di Gödel gli permise di rappresentare la formula metamatematica "il calcolo è coerente" attraverso una formula aritmetica per poi mostrare come questa non possa essere dimostrata utilizzando solo le regole logiche proprie del calcolo (e quindi interne al sistema). Per far ciò l'autore si avvalse di un sistema di numerazione appositamente ideato.

Entrando nello specifico di tale rappresentazione (attraverso la "numerazione" di Gödel) si decide di assegnare ad ogni segno, variabile e quantificatore un numero.

Ad esempio:

$$\neg \ numero\ di\ Godel\ 1$$
$$\vee \ numero\ di\ Godel\ 2$$
$$\supset \ numero\ di\ Godel\ 3$$
$$x \ numero\ di\ Godel\ 11$$
$$y \ numero\ di\ Godel\ 13$$
$$p \ numero\ di\ Godel\ 11^2$$
$$Q \ numero\ di\ Godel\ 13^3$$

....

Per semplificare ulteriormente conveniamo di associare ad una formula un numero unico. Questo è il prodotto di un elenco di numeri primi (tanti quanti sono i segni della formula – disposti in ordine crescente a partire da 2) ognuno di questi elevato al numero di Gödel del segno stesso.

La formula $(\exists x)(x = sy)$ (vi è almeno un x tale che questo è il successore immediato di y) viene tradotta in:

$$2^8 \times 3^4 \times 5^{11} \times 7^9 \times 11^8 \times 13^{11} \times 17^5 \times 19^7 \times 23^{13} \times 29^9$$

Come detto gli apici della sequenza di numeri primi sono il numero di Gödel in cui è tradotto ogni simbolo della formula di partenza che quindi può essere riassunta in un unico numero m.

L'operazione è reversibile e a partire dal numero m è possibile (per tutti i numeri m – non tutti i numeri interi sono numeri di Gödel) ricostruire la formula originaria.

Inoltre se $(\exists x)(x = sy) = m$ e se $(\exists x)(x = s0) = n$ possiamo anche riassumere il sistema di formule con il numero $k = 2^m \times 3^n$. Insomma in un singolo numero possiamo condensare un segno elementare, una sequenza di segni, una sequenza di sequenze di segni. Tale soluzione consente di aritmetizzare il calcolo formale.

Il passo successivo consiste nel mostrare come le rappresentazioni metamatematiche intorno alla proprietà delle espressioni possono essere rappresentate nel calcolo stesso: se a ogni espressione del calcolo è associato ad un numero, allora una proposizione metamatematica intorno alle espressioni può essere costituita da una proposizione attorno ai numeri (di Gödel) e le loro relazioni aritmetiche. Lo studio delle relazioni metamatematiche diviene quindi lo studio delle relazioni tra numeri interni cosicché tutto è stato aritmetizzato.

La dimostrazione procede quindi attraverso i seguenti passi:

- Gödel mostrò come sia possibile la creazione di una formula aritmetica che rappresenti l'espressione metamatematica "$la\ formula\ G\ non$ è $dimostrabile$";
- La formula G è dimostrabile solo se lo è anche la sua negazione $\neg G$. Se quindi sia G che $\neg G$ sono dimostrabili dagli assiomi il calcolo risulta essere contraddittorio il che, per contro, vuol dire che se gli assiomi sono coerenti è impossibile dimostrare sia G che $\neg G$ che quindi risultano essere indecidibili;
- Si dimostra che G pur non essendo dimostrabile è vera;
- Visto che G è insieme vera ed indimostrabile, gli assiomi sono incompleti e non tutte le verità possono essere da essi desunte;
- Si costruisce una formula aritmetica A che rappresenti la proposizione metamatematica '$l'aritmetica$ è $coerente$' e si dimostra che la formula $A \supset G$ è

formalmente dimostrabile. Egli dimostrò infine che A non è dimostrabile. Ne segue che la coerenza dell'aritmetica non può essere stabilita con argomenti rappresentati nel calcolo formale aritmetico.

Presentiamo una versione riassunta di come i passi logici dimostrativi sopra indicati vengono sviluppati.

Partiamo dalla formula: $(x)\neg Dim(x,z)$. Se $Dim(x,z)$ indica l'esistenza di una relazione aritmetica tra le variabili x e z in cui x è il numero che fa riferimento alle formule della dimostrazione di z e z è la formula dimostrata. Ne consegue che l'espressione

$$(1) \qquad (x)\neg Dim(x,z)$$

equivale a dire 'per ogni x la sequenza di formule con numero di Goedel x non è una dimostrazione della formula con numero di Gödel z', ovvero la formula con numero di Gödel z non è dimostrabile.

Per dimostrare questo l'autore utilizzò un caso particolare della formula sopra indicata:

$$(2) \qquad (x)\neg Dim[x, sost(y,13,y)]$$

In quest'espressione $sost(y,13,y)$ designa un numero e precisamente il numero di Gödel della formula che si ottiene dalla formula con numero di Gödel y, sostituendo alla variabile con un numero di Gödel 13 il numerale corrispondente a y. La formula $(x)\neg Dim[x, sost(y,13,y)]$ dice quindi che 'la formula il cui numero di Gödel è il numero della formula, ottenuta dalla formula con numero di Gödel y, sostituendo alla variabile con numero di Godel 13 il numero corrispondente a y, non è dimostrabile'.

Continuando, la formula $(x)\neg Dim[x, sost(y,13,y)]$ possiede un numero di Godel che possiamo chiamare n. Sostituendo n a y otteniamo$(x)\neg Dim[x, sost(n,13,n)]$. Il numero di Godel di questa formula è $sost(n,13,n)$.

Abbiamo ottenuto quindi una formula

$$(3) \qquad (x)\neg Dim[x, sost(y,13,y)]$$

che asserisce di se stessa di non esser dimostrabile (visto che $sost(n, 13, n)$ è il numero di Gödel della formula stessa).

L'argomentazione prosegue mostrando come se la formula (3) fosse dimostrabile lo sarebbe anche la sua negazione e quindi ne conseguirebbe l'incoerenza degli assiomi. Se si assume, come fa Gödel, che gli assiomi sono quindi coerenti, la formula (3) (e la sua negazione) non possono che essere *indecidibili*.

Infine viene dimostrato che $(x)\neg Dim[x, sost(n, 13, n)]$ enuncia una proprietà numerica definita e vera. Di conseguenza, visto che la metamatematica è stata rappresentata nell'aritmetica ed avendo dimostrato la formula essere vera, ne consegue che è vero anche il relativo significato metamatematico che afferma l'indimostrabilità della formula stessa che quindi è vera ma indimostrabile.

Tale conclusione certifica l'incompletezza del sistema assiomatico di partenza e l'esistenza quindi di verità che non possono essere dedotte dagli assiomi (ovviamente sempre assumendo la coerenza degli assiomi).

Tutti i passaggi sopra brevemente riassunti ci portano quindi a poter dire:

se l'aritmetica è coerente, essa è incompleta

Possiamo dire inoltre: 'vi è almeno una formula aritmetica che non è dimostrabile', ovvero 'vi è almeno un numero y tale che, per ogni numero x, x non sta nella relazione Dim con y':

$$(\exists y)(x)\neg Dim(x, y)$$

Se a questa annotazione aggiungiamo la formula che certifica l'esistenza di verità indimostrabili (incompletezza del sistema) $(x)\neg Dim[x, sost(n, 13, n)]$ giungiamo alla seguente:

$$(4) \qquad (\exists y)(x)\neg Dim(x, y) \supset (x)\neg Dim[x, sost(n, 13, n)]$$

Ed indicando antecedente e successore della relazione con A e G la (4) può essere scritta come:

$$A \supset G$$

L'antecedente A non può essere dimostrabile altrimenti, per la regola logica di separazione, lo sarebbe anche G, cosa che abbiamo visto non essere possibile. Ancora una volta, se l'aritmetica è coerente, la formula A non è dimostrabile.

Stante la coerenza dell’aritmetica, la completezza del sistema dell’aritmetica non può essere provata da un ragionamento dotato di rappresentazione nell’ambito del formalismo dell’aritmetica.

In conclusione, le dimostrazioni di Gödel confermano l’impossibilità di una dimostrazione finitistica assoluta di coerenza dell’aritmetica rappresentabili all’interno dell’aritmetica; il metodo assiomatico ha dei limiti intrinsechi dal momento che è possibile asserire verità sui numeri nonostante queste verità non possano essere dimostrate a partire dagli assiomi stessi.

2.6 Turing

Alan Turing (Londra, 23 giugno 1912 – Manchester, 7 giugno 1954) analizzò il problema posto da Hilbert del *Entscheidungsproblem* (problema della decisione), ovvero sull’esistenza o meno di un metodo attraverso il quale potesse essere deciso se ogni proposizione matematica fosse dimostrabile o non dimostrabile. Primo scoglio di questo quesito fu la definizione del significato di ”metodo definito” o ”procedura effettiva”. La soluzione avanzata dal matematico inglese è quella che poi sarà chiamata *Macchina di Turing* ovvero una macchina teorica in grado di scomporre i problemi in singole operazioni elementari.

Il formalismo della macchina di Turing è modellato sulla telescrivente, immaginata leggermente più grande per consentire al nastro (potenzialmente infinito) di spostarsi in entrambe le direzioni ed una "testa" che potrebbe leggere, cancellare e stampare nuovi simboli, piuttosto che solo leggere e fare dei fori. La macchina è una macchina teorica nel senso che non era intenzione dell’autore costruirla: fondamentale è la sua ”possibilità” tecnica e quindi la possibilità di ri-creare il formalismo voluto attraverso le operazioni elementari della telescrivente. Scopo di Turing era quindi trasferire su una macchina ”possibile” i processi meccanici eseguiti da un essere umano.

Per facilitare la comprensione immaginiamo di chiedere a questa macchina se un certo numero dato sia pari o dispari. Possiamo creare un algoritmo, un programma per la macchina, così definito:

- quando la macchina parte si trova nello stato Q;
- legge la prima cifra del numero (non sappiamo quanto lungo), lo cancella e si sposa a destra nella cifra successiva passando allo stato E nel caso in cui il numero cancellato sia pari, e passando nello stato O nel caso in cui il numero cancellato sia dispari;
- la macchina cancella tutte le cifre fino ad arrivare alla fine trovando uno spazio vuoto. Ora la macchina si trova nello stato E o O e come da istruzioni stampa il numero 1 nel caso in cui la macchina si trovi in stato E e 0 nel caso sia nello stato O;
- la macchina quindi si ferma fornendoci l'informazione del numero se pari o dispari attraverso l'annotazione del 1 (stato E – pari) o dello 0 (stato O – dispari) sul nastro.

Tale trasposizione di istruzioni su un programma in una macchina "possibile" rende subito evidente come vi possano essere delle istruzioni senza fine. Ad esempio supponiamo le seguenti istruzioni:

- Se la macchina è nello stato Q e legge 1, sostituisce 1 con 1 e passa nella casella a destra restando nello stato Q;
- Se la macchina è nello stato Q e legge 2, sostituisce 2 con 2 e passa nella casella a sinistra restando nello stato Q.

Tali istruzioni possono essere formalizzate in quintuple (comandi con 5 caratteri) come di seguito riportato:

$$Q1{:}\,1 \rightarrow Q$$
$$Q2{:}\,2 \leftarrow Q$$

Ed è evidente che se la macchina affronta il numero 12, questa non finirà mai di lavorare.

Tornando all' *Entscheidungsproblem,* Turing, sulla scorta dell'esperienza di Gödel, propone una codifica sia dei numeri sul nastro che dei simboli utilizzati nelle quintuple/programmi in numeri. Tale soluzione permette di identificare con un numero la serie di istruzioni della macchina.

In alcuni casi la macchina si fermerà, in altri continuerà ad eseguire il proprio processo all'infinito. L'insieme dei numeri del primo tipo (che corrispondono ai programmi per i quali la macchina porta a termine il proprio compito) è definito come il suo *insieme di fermata*. Si può pensare all'insieme di fermata di una macchina di Turing come un *pacchetto* e il codice numerico di quella macchina come l'etichetta di quel pacchetto. Tale situazione consente l'applicazione del metodo della diagonale di Cantor[6] e quindi la costruzione di un insieme di

---

[6] Il metodo della diagonale del matematico tedesco Georg Ferdinand Ludwig Philipp Cantor (San Pietroburgo, 3 marzo 1845 – Halle, 6 gennaio 1918) conferma che non esiste corrispondenza biunivoca tra i numeri naturali e i numeri reali, dove per numeri reali si intendono i numeri ai quali è possibile attribuire uno sviluppo decimale finito o infinito e per numeri naturali quelli usati per contare e ordinare (0,1,2,3,4,...). Per comprendere tale metodo immaginiamo di avere dei pacchi ognuno dei quali reca un'etichetta e con la caratteristica che gli oggetti nei pacchi e le etichette sono dello stesso tipo. Prendiamo ad esempio i semi delle carte francesi ed e usiamo gli stessi per identificare quattro gruppi di semi:

♣↔(♥♦); ♦↔(♥♦); ♥↔(♥♠♦); ♠↔(♣♦)

Tali correlazioni possono essere riportate in una tabella in cui il simbolo + rappresenta la presenza del seme all'interno del gruppo e il – la non presenza. Organizzando quindi nella colonna a sinistra le varie etichette e nella riga in alto i semi presenti o meno nei singoli pacchetti otteniamo

	♣	♦	♥	♠
♣	-	+	+	-
♦	-	+	-	+
♥	-	+	+	+
♠	+	+	-	-

Il metodo della diagonale consiste nel raccogliere oggetti dello stesso genere in un nuovo pacco che però contiene cose diverse dai pacchi già etichettati. Creiamo quindi una nuova tabella in cui per ogni etichetta il segno opposto a quello presente sulla diagonale della precedente tabella:

numeri naturali (che possiamo chiamare D) diverso dall'insieme di fermata di qualsiasi macchina di Turing. « D sarà quindi formato esclusivamente da codici numerici di macchine di Turing, e il codice numerico di una qualsiasi macchina di Turing apparterrà a D se e solo se non appartiene all'insieme di fermata di quella macchina. In altre parole, se il codice numerico di una certa macchina di Turing appartiene al suo insieme di fermata, allora non appartiene a D; e viceversa, se non appartiene all'insieme di fermata, allora appartiene a D. Nell'uno come nell'altro caso D non può essere l'insieme di fermata della macchina in questione; e poiché questo vale per ogni macchina di Turing, concludiamo che D non è l'insieme di fermata di nessuna macchina di Turing»[7].

Tale conclusione risulta fondamentale ai fini del problema dell'*Entscheidungsproblem* posto da Hilbert. Infatti se ipotizziamo l'esistenza di un algoritmo capace di rispondere a tutte le domande matematiche, e che è possibile dimostrare che alcuni problemi sono insolubili, lo stesso

♣	♦	♥	♠
+	-	-	+

Il nuovo pacco sarà quindi costituito dalla coppia ♣♠. Ci chiediamo come sia possibile verificare che la nuova coppia sia diversa dai gruppi precedentemente etichettati. Infatti non può essere uguale a quello dell'etichetta ♣ che non contiene ♣, seme invece presente nel nuovo pacco; non può essere quello con l'etichetta ♦ visto che contiene ♦ a differenza della nuova coppia che non lo contiene, e così per tutti i semi. Ora, se i semi sono insiemi e l'etichetta un modo per stabilire una corrispondenza biunivoca tra insiemi ed elementi, si dimostra che, indipendentemente che l'insieme sia finito o infinito, se si usano gli elementi dell'insieme per etichettare un qualche insieme particolare degli stessi elementi, si potrà sempre ottenere con il metodo della diagonale un nuovo insieme di tali elementi diverso da tutti quelli etichettati, in altre parole vi sono più numeri reali che numeri naturali. Tale risultato apre la strada all'ipotesi che esistano infiniti diversi, alcuni maggiori di altri.

[7] M. Davis, *Il calcolatore universale*, trad. it. G. Rigamonti, Adelphi Edizioni, Milano 2018.

Entscheidungsproblem sarà insolubile e l'insieme D ci fornirà un tale esempio. Si consideri infatti un algoritmo per:

(5) determinare se un dato numero naturale n appartiene o meno a D

Dovrebbe quindi esistere una macchina di Turing, similmente a quando visto per i numeri pari o dispari, capace di eseguire le operazioni necessarie per ottenere la risposta alla (5). Se il numero in ingresso appartiene a D allora correttamente la macchina farà le sue elaborazioni arrestandosi poi e confermando che il numero oggetto di valutazione effettivamente appartiene a D e l'insieme di fermata di questa macchina sarà identico a D. Tale soluzione evidenzia una contraddizione in quanto D è stato costruito usando il metodo della diagonale di Cantor in modo tale che sia diverso dall'insieme di fermata di qualsiasi macchina di Turing.

Infine immaginiamo il numero naturale n che rappresenta sia il numero scritto sul nastro della macchina di Turing che questa macchina deve leggere, sia il numero della macchina stessa. Ipotizziamo che questa macchina n venga avviata e che si fermi dopo una serie di operazioni. Tale conclusione è possibile solo se n non appartiene a D e quindi se avessimo l'algoritmo dell'*Entscheidungsproblem* potremmo usarlo per deciderne l'appartenenza a D, ovvero dato un numero naturale n potremmo testare il presunto algoritmo per controllare se la conclusione deriva dalla premessa. Se così fosse sapremmo che n non appartiene a D ; al contrario sapremmo che non gli appartiene. Ne segue che *Entscheidungsproblem* è algoritmicamente insolubile.

La dimostrazione di Turing parte dal presupposto che è possibile (erano gli anni trenta dello scorso secolo) far eseguire a un'ipotetica macchina un'operazione logica elementare. Il matematico inglese, per supportare la propria tesi ipotizzò una macchina universale ovvero capace di eseguire tutti i calcoli possibili. Tale macchina avrebbe utilizzato il numero sul nastro come il riferimento di una macchina specifica con specifiche istruzioni/quintuple. Egli ideò e dimostrò la possibilità di realizzare una macchina che gestisse quelli che oggi definiremmo programmi. Le scritte sul nastro avrebbero quindi fatto riferimento alle funzionalità di altre macchine che questa sarebbe stata in grado di riprodurre. Così ideata, la macchina universale di Turing può eseguire tutti i calcoli possibili.

Se una macchina di Turing può essere assimilata ad un programma, la macchina di Turing universale è in grado di eseguire infiniti programmi.

Oggi parlando delle macchine di Turing ci viene naturale pensare ai computer; ma questi, nel 1936, semplicemente non esistevano.

3 L’intelligenza artificiale

L’intelligenza artificiale nasce ufficialmente nel 1956 durante una conferenza presso il Dartmount College, Hanover, New Hampshire. Durante questa conferenza viene per la prima volta utilizzato il termine AI ed è presentato il programma LT (Logic Theorism) volto alla creazione di una macchina capace di imitare le capacità umane.

Lo studio procede partendo dall’assunto di base che ogni aspetto della conoscenza e ogni altra caratteristica dell’intelligenza possa essere descritta così precisamente e che sia quindi possibile creare (almeno in linea di principio) una macchina capace di simularla[8].

Se il termine AI comparve ufficialmente nella conferenza del 1956 (e nei documenti preparatori del 1955), già negli anni precedenti alcuni scienziati si erano posti il problema sui limiti teorici delle macchine.

Dopo il contributo allo sforzo bellico, Alan Turing[9] lavora per il "National Physical Laboratory" (NPL) allo sviluppo all'"Automatic Computing Engine" (ACE), uno dei primi tentativi di creare un vero computer digitale. Fu in questo periodo che inizia a esplorare la relazione tra i computer e la natura. Turing, infatti, era dell'idea che si potessero creare macchine che fossero capaci di simulare i processi del cervello

---

[8] J. McCarthy, M. L. Minsky, N. Rochester, C.E. Shannon, *A proposal for the Darmouth summer research project on artificial intelligence*, 1955.

[9] Nel 1940 sempre Alan Turing, a 28 anni, era a capo del gruppo di ricercatori impegnati nella decrittazione delle macchine usate dalla marina tedesca, fra le quali Enigma (sistema di criptografia tedesco caratterizzato dal fatto che i codici cambiavano sempre). La macchina realizzata da Touring e dai sui collaboratori era in grado di decifrare velocemente e correttamente i messaggi dei tedeschi.

umano, nella convinzione che non ci sia nulla, in teoria, che un cervello artificiale non possa fare rispetto al cervello umano.
Nel paper del 1950, *Mind*, il matematico inglese si chiede *"Can a machine think?"*.
A questa domanda se ne affianca una seconda: "Può una macchina essere linguisticamente indistinguibile da un essere umano?". Quest'ultima questione rappresenta il Turing Test (abbreviato con TT). In questo test una macchina e un umano sono rinchiusi in due stanze isolate. Un giudice umano invia domande nelle singole stanze e ne legge le risposte indicando in quale stanza si trovano la macchina e la persona. Se il giudice non è in grado di rispondere correttamente, in quale stanza di trova l'umano oltre il 50% dei casi (che equivale a dire che non si distinguono i due risponditori) allora si può dire che la macchina ha superato il TT.
Ad oggi nessuna macchina ha mai superato questa verifica (lo stesso Turing aveva predetto che entro il 2020 una macchina sarebbe stata in grado di superare la questa prova).

Negli ultimi anni alla versione classica del TT si sono aggiunte numerose varianti quali il test di *Turing Inverso*, il *TT dell'esperto*, il *Test del minimo segnale intelligente*. Tra questi curioso è l'*Ebert Test*. Roger Joseph Ebert (Urbana, 18 giugno 1942 – Chicago, 4 aprile 2013) era un famoso critico cinematografico statunitense; a causa di gravi problemi di salute aveva perso la possibilità di parlare dovendo accontentarsi di una voce sintetizzata dal computer. Visto il proprio lavoro inoltre era stata realizzata una "sua" voce utilizzando le molte ore di registrazione accumulate durante i molti anni da critico. Nonostante questa personalizzazione, la voce sintetizzata non riusciva ad esprimere tutte le sfumature, le intonazioni, le inflessioni ed i tempi della normale voce umana. Una voce artificiale che superi l'*Ebert Test* dovrebbe essere in grado di riprodurre anche queste caratteristiche e riuscire a far ridere le persone raccontando una barzelletta.

Dal punto di vista filosofico la questione dell'intelligenza artificiale era già stata trattata secoli prima dei lavori di Turing. Tra i primi ad interrogarsi sull'argomento possiamo indicare Cartesio[10] che nel 1637 nella quinta parte del "Discorso sul metodo" scrive:

"Qui in particolare mi ero fermato per far vedere che se ci fossero macchine con organi e forma di scimmia o di qualche altro animale privo di ragione, non avremmo nessun mezzo per accorgerci che non sono in tutto uguali a questi animali; mentre se ce ne fossero di somiglianti ai nostri corpi e capaci di imitare le nostre azioni per quanto è di fatto possibile, ci resterebbero sempre due mezzi sicurissimi per riconoscere che, non per questo, sono uomini veri. In primo luogo, non potrebbero mai usare parole o altri segni combinandoli come facciamo noi per comunicare agli altri i nostri pensieri. Perché si può ben concepire che una macchina sia fatta in modo tale da proferire parole, e ne proferisca anzi in relazione a movimenti corporei che provochino qualche cambiamento nei suoi organi; che chieda, ad esempio, che cosa si vuole da lei se la si tocca in qualche punto, o se si tocca in un altro gridi che le si fa male e così via; ma non si può immaginare che possa combinarle in modi diversi per rispondere al senso di tutto quel che si dice in sua presenza, come possono fare gli uomini, anche i più ottusi".

Com'è evidente il test di Cartesio è ancor più severo del TT.

Per una corretta definizione dell'AI è necessario precisare i significati del termine intelligenza. Infatti questo varia a seconda che si miri a replicare un pensiero perfettamente razionale o umano, o si abbia come obiettivo l'imitazione dell'agire razionale o umano. Le possibili risposte sono qui riassunte:

---

[10] René Descartes (La Haye en Touraine - oggi Descartes, 31 marzo 1596 – Stoccolma, 11 febbraio 1650).

L'AI è un campo di ricerca volto a costruire

	Razionalità Umana	Razionalità ideale
Ragionamento	Un sistema che pensa come gli uomini	Un sistema che pensa razionalmente
Azione	Un sistema che agisce come gli uomini	Un sistema che agisce razionalmente

Per Stuart Russell[11] l'intelligenza equivale alla razionalità e scopo dell'AI è la creazione di agenti intelligenti. Per quest'autore, l'intelligenza/razionalità ha come scopo la massimizzazione dell'utilità dell'agente F nell'ambiente E considerando la sequenza degli accadimenti (nell'ambiente) derivanti dall'azione dell'agente stesso. Se definiamo U la valutazione della performance di questi accadimenti, $V(f, E, U)$ rappresenta l'utilità di U dell'agente F nell'ambiente E.

Il comportamento dell'agente perfettamente razionale, volto quindi alla massimizzazione della sua utilità attesa, è quindi formalizzabile dalla seguente equazione:

$$f_{opt} = \underset{f}{\text{arg max}}\ \mathrm{V}(f, E, U)$$

Russell nota come simili algoritmi sarebbero in grado di trovare la soluzione perfetta ed un programma così realizzato sarebbe, ad esempio in una partita a scacchi, imbattibile. Ma, visti i limiti dei programmi si è costretti a limitare la ricerca della soluzione perfetta interrompendo l'elaborazione. Infatti il programma è svolto da un computer con capacità limitate. Se Agent(P, M) è la funzione dell'agente implementata dal programma P nella macchina M e P(M) rappresenta tutti i possibili programmi eseguibili dalla macchina M, il miglior

[11] S. Russell, *Rationality and intelligence*, 1997, Computer Science Division, University of California, Berkeley.

programma limitato[12] (altrimenti la soluzione del problema richiederebbe troppo tempo) è rappresentato dalla seguente equazione

$$P_{ptm,M} = \underset{P \in P(M)}{\arg\max} V\ (Agent(P, M), E, U)\ (f, E, U)$$

In altre parole, per S. Russell l'intelligenza artificiale è la ricerca di programmi volti alla miglior soluzione ed implementati in macchine con limiti temporali e dimensionali.

Si nota ancora come E e U siano variabili molto varie e dipendano, ad esempio, dal gioco che stiamo giocando, sia questo gli scacchi, Jeopardy o Go. Nel caso di un'intelligenza artificiale generica non avremo quindi una sola coppia di E ed U. Infatti un tale tipo di intelligenza dovrebbe includere tutte le variabili possibili e non limitarsi solo a specifiche coppie, costruite per specifiche funzioni.

3.1 Searle e l'argomento della stanza Cinese

Nel 1980 il filosofo americano John Searle (Denver, 31 luglio 1932) propose un celebre esperimento mentale noto come l'argomento della stanza cinese volto ad evidenziare, secondo l'autore, i limiti dei computer. È utile far precedere l'argomento di Searle da altri esempi storici.

Già Leibniz nella sezione 17 del suo "Monadologie" aveva scritto:

"D'altra parte siamo costretti a confessare che la percezione, e ciò che ne dipende, è inesplicabile per ragioni meccaniche, cioè per figure o per movimenti: e se immaginiamo una macchina, la cui struttura faccia pensare, sentire, percepire, si potrà concepire ingrandita conservando le medesime proporzioni, di maniera che vi si possa entrare come in un molino. Ciò posto, se ci troveremo a

[12] Si tratta di una limitatezza decisa da un programmatore che sceglie di interrompere il programma quando ritiene l'elaborazione sia sufficiente e che non valga la pena investire altre risorse per migliorare il risultato.

visitarlo al di dentro, vi troveremo dei pezzi che si sospingono gli uni e gli altri, ma non vi troveremo giammai come spiegare una percezione: la quale perciò nella sostanza semplice, e non nel composto o nella macchina è mestieri cercarla. Così non vi è che questo, che si possa trovare nella sostanza semplice, vale a dire le percezioni e loro cambiamenti. Dunque in ciò solo possono consistere tutte le azioni interne delle sostanze semplici"[13].

Il "programma di carta" dello stesso Turing. Lo scienziato inglese disse di aver scritto un programma (chiamato *Turochamp*, fusione tra il nome dello scienziato e quello di David Champernowne, collega con il quale condivise il tentativo) per il gioco degli scacchi. Si trattava di una serie di istruzioni in linguaggio naturale che permettevano ad un giocatore, ignorante del gioco degli scacchi, di giocarci. Una qualsiasi persona quindi poteva, seguendo passo passo le istruzioni sfidare un giocatore vero. Anche se all'apparenza la persona che segue le istruzioni sembra comprendere il gioco stesso, nella realtà non sa nulla del gioco degli scacchi ma banalmente segue regole codificate. La prima partita tra un uomo ed una macchina (in realtà un altro uomo che eseguiva pedissequamente il "programma di carta" del matematico) fu giocata nel 1952; il "giocatore uomo" vince facilmente in 29 mosse.

Nel 1980 John Searle pubblica *"Minds, Brains and Programs"* all'interno della rivista *The Behavioral and Brian Sciences*, il suo celebre argomento della stanza cinese.
L'autore immagina una stanza con all'interno un uomo che nulla sa di cinese. Nella stanza vengono introdotti degli ideogrammi che ovviamente l'uomo all'interno non sa leggere. Ma questi ha a disposizione un enorme dizionario e regole che gli permettono di fornire delle "risposte" ai biglietti che riceve.

---

[13] G. Leibniz , *La Monadologia* (1720) trad. dal francese di Marianna Bacinetti (1856).

Vengono quindi inviati all'esterno dei nuovi biglietti che sono le "risposte" a quelli ricevuti. Il parallelismo con i computer è di facile intuizione. Il calcolatore elabora gli input servendosi di un database (dei dizionari) e fornendo quindi degli output corretti (ovviamente se il dizionario è corretto). La macchina supererebbe il test di Touring ma come l'uomo all'interno della stanza, non si potrà certo dire, secondo Searle, che conosca il cinese.

Una prima risposta a questo argomento è la così detta "risposta di sistema". In questa si concede che l'individuo nella stanza non conosce il cinese ma si aggiunge che questi non è che una parte del sistema che include gli input e il dizionario: il sistema nel suo complesso conosce il cinese. La risposta di Searle non concede spazio a questa proposta affermando che il singolo uomo potrebbe anche imparare tutto a memoria e girovagare per la Cina conversando in cinese. Ma ciò non avrebbe modificato affatto la sua non conoscenza del cinese. Infatti una correttezza delle regole sintattiche non comporta il salto necessario per la vera conoscenza semantica della lingua.

Una seconda obiezione è quelle definita "della mente virtuale". Anche qui si concede che la persona all'interno della stanza non conosce il cinese ma si afferma che esiste una mente virtuale che va oltre gli aspetti fisici (stanza, persona, dizionario) ed è il sistema così concepito che conosce il cinese. Se potessimo porgli la domanda "quando hai imparato il cinese?" questo risponderebbe e tale output sarebbe la risposta del sistema e non dell'uomo nella stanza.

La "risposta del robot" accetta la convinzione di Searle che l'uomo nella stanza effettivamente non conosce la lingua, ma in questo punto di vista si sostiene che invece le cose cambiano e se al calcolatore viene aggiunto un "corpo" con sensori e telecamere e se questa creatura avesse la liberta di girare per la stanza. Come i bambini apprenderebbe; questo "essere" potrebbe davvero conoscere il cinese. Per Searle, anche questa riposta è inappropriata in quanto l'aggiunta di nuove informazioni non è sufficiente a quel salto qualitativo tra sintassi e semantica.

La risposta "del simulatore della mente" propone di concepire un computer che replichi esattamente tutte le connessioni neurali di un cinese madrelingua. La risposta di Searle è nella sostanza uguale a quella precedente.

Alla base dell'argomento della stanza cinese e della conclusione (per Searle) circa l'impossibilità dei computer di comprendere (nel senso umano del termine) sta la considerazione che i calcolatori hanno la capacità di elaborare le informazioni ricevute solo in maniera sintattica ovvero attraverso la mera combinazione programmata dei simboli in input. Infatti rispondono solo a sequenze di simboli ed operazioni non potendo cogliere il significato degli stessi. Gli uomini, invece, rispondono utilizzando il significato, la semantica, di quegli stessi simboli. Searle riassume così questo passaggio chiave: "*Syntax is not by itself sufficient for, nor constructive of, semantics*". Tale posizione si esplica poi nella dimostrazione di questo stesso autore sull'impossibilità di un'intelligenza artificiale forte:

1. i programmi sono solamente formali (sintassi);
2. le menti umane hanno contenuti mentali (semantica);
3. la sintassi di per sé stessa non è né costitutiva né sufficiente per un contenuto semantico;
4. ne consegue che i programmi non sono di per sé né costruttivi, né sufficienti, per una mente.

Ancora, i simboli, per definizione non hanno significato e quindi senza un aiuto esterno non è possibile passare dalla sintassi alla semantica.

Altro corollario della tesi sostenuta da Searle è che il calcolo computazionale non può portare alla coscienza. L'autore, infatti, sostiene che quest'ultima ha caratteri intrinsecamente biologici.

Egli aggiunge inoltre che "la computazione esiste solo in relazione a qualche agente o osservatore che decide per una interpretazione computazionale del fenomeno".

Un altro punto di vista interessante per guardare all'argomento della stanza cinese è quello della simulazione o duplicazione. Certamente nessuno confonderebbe, secondo il filosofo, un programma per la simulazione del meteo dal meteo stesso. Similmente sarebbe un errore confondere un computer che simula la conoscenza dalla conoscenza vera. Più a fondo in questo ragionamento, emergono alcuni problemi: un cuore artificiale è un duplicato di un cuore o un vero cuore solo fatto con materiali diversi dal solito?
Searle sostiene che «il cervello è un organo biologico specifico e le sue proprietà biochimiche gli consentono di causare la coscienza e altri tipi di fenomeni mentali»[14] e che l'imitazione dell'intelligenza nulla ha a che fare con l'intelligenza stessa. Continua Searle: «a meno che non avvenga un miracolo, non … potremmo digerire una pizza eseguendo il programma che simula tale digestione»[15]. Tale argomentazione non convince a pieno in quanto teoricamente – ad esempio – le funzioni digestive possono essere replicate artificialmente (o con tessuti creati artificialmente). Inoltre non son esplicitate le caratteristiche biologiche che renderebbero impossibile la replicabilità digitale della ragione umana, non solo come imitazione della relazione input-output, ma come autentica capacità intellettiva.

Un ulteriore riflessione riguarda l'evoluzione e nello specifico il fatto che secondo la teoria evoluzionista darwiniana, statisticamente sopravvivono, e quindi hanno maggior possibilità di trasferire il proprio corredo genetico, gli individui che si adattano meglio all'ambiente. Ora, se un sistema, pur non comprendendo i significati, offre le migliori risposte, questo sarà preferito nella selezione naturale e nulla conta che la comprensione sia reale o meno. L'intima comprensione non è quindi elemento essenziale alla selezione naturale: è sufficiente l'adattamento

[14] J. Searle, *La mente è un programma?,* tratto dalla rivista Le Scienze n. 259, marzo 1990, pag. 19.

[15] Ibidem.

all'ambiente, anche se inconsapevole. Inoltre l'uomo, specie cosciente e consapevole di aver superato (ad oggi) la prova della selezione naturale, indica questa vittoria come la conferma della propria superiorità considerando la propria coscienza come l'unica vera.

3.2 La nazione cinese

Ned Block (1942 - Chicago) presenta, in contrapposizione alla teoria funzionalista[16], l'esperimento mentale della "nazione Cinese". Immaginiamo che tutta la popolazione dello stato più popoloso al mondo sia organizzata in modo tale da replicare le funzionalità del cervello. In questo esperimento ogni individuo rappresenta un singolo neurone che si mette in comunicazione con altri individui seguendo regole precise. Per il funzionalismo l'insieme dei cinesi così organizzati è una gigantesca mente con stati mentali propri quali coscienza e dolore; nessuna delle sue parti, nel caso gli individui, prova queste sensazioni.

3.3 Putnam e i cervelli nella vasca

Hilary Putnam (Chicago, 31 luglio 1926 – Arlington, 13 marzo 2016) rifiuta sia il realismo metafisico, sia lo scetticismo. Secondo l'autore, sia chi cerca una realtà metafisica sia lo scettico che nega tale possibilità, si pongono nella posizione di

[16]La teoria funzionalista afferma che gli stati mentali (desideri, convinzioni, ecc.) siano costituiti solamente dal loro ruolo, cioè dalla loro funzione, la loro relazione causale, rispetto ad altri stati mentali, percezioni e comportamento. Giacché stati mentali possono essere definiti in base al loro ruolo funzionale, essi sono molteplicemente realizzabili; in altre parole, possono manifestarsi in vari sistemi, anche artificiali (ad es. calcolatori), se il sistema computa le funzioni adatte. L'analogia mente/computer, che vede il cervello paragonato all'hardware e la mente al software, costituisce l'emblema di gran parte delle teorie funzionaliste della mente.

un osservatore privilegiato, inesistente, esterno ai nostri schemi descrittivo-concettuali. Assumono infatti l'esistenza di un mondo precostituito e a noi indipendente. Il primo ritiene questo sia conoscibile, per il secondo ci è per sempre negato. Per Putnam, in continuità con la tradizione kantiana, non ha senso porsi all'esterno dei nostri schemi concettuali, dei nostri linguaggi.

L'esperimento dei cervelli in una vasca rappresenta la sua più nota critica al realismo metafisico. In questo esperimento si immagina un mondo in cui tutti i cervelli rimossi dai corpi siano posti all'interno di una vasca, che vi sia un sistema per alimentarli e che tutte le terminazioni nervose e i neuroni siano collegati ad un avanzatissimo super-computer in grado di rispondere agli impulsi come se i cervelli fossero normalmente ognuno nel proprio corpo. Il super-computer sarebbe in grado di fornire la medesima esperienza della normale vita quotidiana senza però che vi sia alcuna relazione con eventi e oggetti del mondo concreto perché, appunto, i cervelli sono solo in una vasca.

«[...] Sembra che ci siano persone, oggetti, il cielo ecc., ma in realtà l'esperienza della persona (la vostra esperienza) è in tutto e per tutto il risultato degli impulsi elettronici che viaggiano dal computer alle terminazioni nervose. Il computer è così abile che se la persona cerca di alzare il braccio la risposta del computer farà sì che "veda" e "senta" il braccio che si alza. Inoltre, variando il programma lo scienziato malvagio può far sì che la vittima "esperisca" qualsiasi situazione o ambiente lo scienziato voglia. Può anche offuscare il ricordo dell'operazione al cervello, in modo che la vittima abbia l'impressione di essere sempre stata in quell'ambiente. [...] magari l'universo [...] consiste solo di macchinari automatici che badano a una tinozza piena di cervelli. Supponiamo che il macchinario automatico sia programmato per dare a tutti noi un'allucinazione collettiva [...] Quando sembra a me di star parlando a voi, sembra a voi di star ascoltando le mie parole. Naturalmente le mie parole non giungono per davvero alle vostre orecchie, dato che non avete (vere) orecchie, né io ho una vera bocca e una vera lingua. Invece, quando produco le mie parole quel che succede è che gli impulsi efferenti viaggiano dal mio cervello al computer, che fa sì che io "senta" la mia stessa voce che dice quelle parole e "senta" la lingua muoversi, ecc., e anche che voi "udiate" le mie parole, mi "vediate" parlare, ecc. In questo caso, in un certo senso io e voi siamo davvero in comunicazione. Io non mi inganno sulla vostra esistenza reale, ma solo sull'esistenza del vostro corpo e del mondo esterno, cervelli esclusi»[17].

Lo scettico direbbe che il cervello nella vasca non può decidere se effettivamente si trova nella vasca o meno in quanto tutte le esperienze che vive, gli stimoli che riceve, sono esattamente gli stessi sia che si trovi, appunto, nella vasca che nel mondo reale. Il proprio ambiente quindi non pare creato dal super-computer ed è del tutto identico all'ambiente sperimentato da un normale cervello attaccato ad un corpo. Lo scettico sostiene che non siamo in grado di stabilire se le nostre credenze corrispondano a verità e di conseguenza non sappiamo se le nostre asserzioni parlano effettivamente di noi e del mondo esterno.

Putnam osserva che un cervello da sempre in una vasca non potrebbe pensare e parlare della sua condizione e, nel caso dicesse "io sono un cervello in una vasca", direbbe necessariamente il falso.

Infatti, banalmente, nel caso questo cervello vivesse nel mondo reale, l'affermazione sarebbe ovviamente falsa.

Nell'altro caso in cui effettivamente il cervello si trovasse nella vasca la sua affermazione potrebbe essere tradotta con "io sono ciò che i miei stimoli nervosi mi hanno convinto che è un 'cervello', e vivo in un'immagine che essi mi hanno convinto a chiamare 'vasca'", ovvero "io sono un cervello-immagine in una vasca-immagine". In altre parole, il cervello nella vasca, poiché non può far riferimento a ciò a cui noi (che siamo cervelli normali) facciamo riferimento, non potrà pensare a cervelli o vasche reali ma alle immagini di questi ricreate nel mondo virtuale.

[17] Hilary Putnam, *Brains in a Vat*, 1981, pp.6-7.

Essendo quindi preclusi i riferimenti al mondo reale, il cervello nella vasca non potrà emettere alcuna affermazione o negazione del mondo esterno in quanto non né ha mai avuto esperienza. Quindi se questo cervello dicesse "io sono un cervello in una vasca" direbbe il falso in quanto questo sarebbe in effetti in una vasca ma le sue parole non possono prendere a riferimento la vasca reale ma necessariamente si riferiscono alla vasca-immagine di cui ha la sola esperienza.
Tale argomentazione poggia sulla teoria del riferimento che "ci vieta di credere al fatto che determinate rappresentazioni mentali si riferiscano *in ogni caso* a specifiche cose esterne la cui configurazione è del tutto indipendente dalla nostra mente"[18].
I riferimenti usati dei cervelli nella vasca sono infatti diversi da quelli usati dai cervelli normali cosicché quanto dicono i primi hanno un significato diverso: "dire sono un cervello in una vasca" per un cervello in una vasca ha un altro significato (ad di là della veridicità della proposizione) rispetto alla stessa frase detta da un cervello normale.
L'esperimento mentale si auto-confuta ed è quindi inconsistente. La distanza tra mondo reale e mondo visto dall'uomo, distanza incolmabile sia per lo scettico che per il realista metafisico, non ha senso.

3.4 AI forte e AI debole

L'intelligenza artificiale (acronimo AI), sin dalla sua origine, ha come obiettivo la realizzazione di macchine che imitino l'agire o il pensiero umano. L'AI debole mira a realizzare macchine che agiscano come gli uomini. L'AI forte, invece, si pone l'obiettivo ben più ambizioso ipotizzando che un sistema artificiale possa replicare il pensiero.

---

[18] Sacchetto, M. (2002), "Putnam: logica, filosofia del linguaggio e riflessione etica", in G. Fornero e S. Tassinari (a cura di), *le filosofie del Novecento*, Milano, Bruno Mondadori.

La dimensione di quest’ultima impresa viene efficacemente rappresentata dal discorso del 1949 del professore inglese Geoffrey Jefferson (Stockton-on-Tees, County Durham, 10 April 1886 - Manchester, 29 January 1961): «solo quando una macchina avrà scritto un sonetto o composto un concerto in virtù dei pensieri e delle emozioni provate, e non per la disposizione casuale dei simboli, potremo essere d’accordo sul fatto che eguaglia il cervello: non deve solo aver scritto, ma anche aver la conoscenza di aver scritto».
La coscienza svolge evidentemente un ruolo fondamentale ma, sottolinea Turing, è importante porre la questione in modo corretto. Infatti, noi non abbiamo alcuna prova del pensiero degli altri esseri umani e quindi sarebbe ingiusto pretendere una prova dalle macchine. Noi non abbiamo alcuna prova del pensiero altrui ma semplicemente supponiamo che dietro le loro azioni vi sia un pensiero cosciente. E perché mai non dovremmo estendere questa buona fede anche alle macchine?

Un’importante critica alle pretese dell’AI forte è di tipo matematico. Nel 1931 Gödel pubblicò i propri teoremi sull’incompletezza nei quali si dimostra che in un sistema assiomatico formale vi sono verità indimostrabili. Tali teoremi sono utilizzati da chi sostiene che i computer sono strutturalmente inferiori all’uomo essendo questi delle macchine formali. Tale obiezione può essere respinta in quanto il teorema di incompletezza di Gödel si applica solo ai sistemi formali abbastanza potenti da contenere l’aritmetica. Questo include le macchine di Turing ma non tutti i computer sono macchine di Turing. Le macchine di Turing sono infinite mentre i computer hanno dimensione finita e ogni computer può esser definito come un sistema (molto grande) di logica proposizionale, non soggetto al teorema di incompletezza.
Un'altra obiezione a chi usa i teoremi del logico austriaco per sostenere la superiorità della mente è la seguente: “anche se accettiamo che i computer siano limitati in ciò che possono dimostrare, nulla prova che gli esseri umani siano

immuni da tali limitazioni."[19]. All'affermazione, infatti, che un sistema formale è impossibilitato a fare x, non segue la spiegazione di come l'uomo sia in grado di fare quella stessa cosa attraverso un metodo informale. Ed una tale dimostrazione è per definizione impossibile in quanto sarebbe necessario formalizzare questo metodo informale, cosa appunto che non si può dare. "Tutto quel che resta quindi è un appello all'intuizione che gli esseri umani possano in qualche modo eseguire azioni sovrumane di intuizione matematica"[20].

La posizione "dell'informalità del comportamento" sostiene che le macchine non possano, attraverso un insieme di regole, imitare il comportamento umano, considerato troppo complesso. Uno dei sostenitori di questa tesi è stato il filosofo Hubert Dreyfus (Terre Haute, 15 ottobre 1929 – Berkeley, 22 aprile 2017) che difendeva le competenze umane in quanto inserire all'interno di un contesto olistico e di cornice. Ne sarebbe un esempio il grande maestro degli scacchi che non ha bisogno di pensare alle mosse in quanto queste gli si presentano nella mente. A tale argomentazione si può rispondere che nessuno ha spiegato come e perché gli si presentano nella mente e quindi, ancora una volta, non viene data ragione di questa ipotizzata superiorità umana.
E' indubbio però che gli uomini agiscono in un corpo e in un ambiente che rappresenta un fattore necessario e importantissimo per lo sviluppo delle competenze umane. Similmente, per le intelligenze artificiali o per la ricerca delle stesse, l'uso di sensori e processi cognitivi robotici sono elementi fondamentali.

Altro argomento relativo alla possibilità o meno dell'AI forte è quello relativo al dualismo mente-corpo ed al fatto che qualcuno potrebbe obiettare che gli uomini hanno menti reali, mentre nelle macchine questo non si sa. Già Cartesio

[19] S. Russell, P. Norvig, Intelligenza artificiale, Pearson Italia, Milano – Torino 2010- pag. 571
[20] Ibidem.

nel *Meditationes de prima philosophia* scindeva mente da corpo individuandoli rispettivamente in *res cogitans* e *res extensa*. Tale distinzione attribuisce alla mente un processo privo di qualsiasi estensione materiale e di contro una dimensione spaziale per la materia. Essendo mente e corpo, *res cogitans* e *res extensa*, due elementi completamente separati, la teoria pone il problema di come la mente possa influenzare e "comandare" il corpo se non vi sono punti di contatto. A questa proposta si oppone la teoria monista, o del fisicalismo, che sostiene che la mente non è separata dal corpo e che gli stati mentali sono stati fisici. Tale posizione concede, in linea di principio, la possibilità di un AI forte, e l'obiettivo dei fisicalisti è di spiegare come gli stati fisici (le configurazioni di molecole e i processi elettrochimici) del cervello siano allo stesso tempo stati mentali, come ad esempio provare dolore o mangiare un hamburger. Alcuni filosofi fisicalisti dovettero spiegare tale possibilità per poter dire con precisione cosa significa per una persona – e per estensione, per una macchina – essere in un determinato stato mentale. Per il fisicalismo quindi la descrizione di uno stato mentale di una persona è determinata dal suo stato cerebrale. Quindi se desidero mangiare un hamburger avrò una specifica configurazione cerebrale, diversa nel caso desiderassi mangiare un qualcos'altro. Tale posizione è messa in discussione dall'esperimento mentale dei "cervelli in una vasca" di Puntam. Se il cervello è in realtà immerso in una vasca, allora lo stato mentale di quel cervello scollegato dal mondo che afferma di "mangiare un hamburger" sarà diverso (non ha mai mangiato un vero hamburger) dallo stato mentale di un altro cervello che ha mangiato un hamburger. L'esempio sembra quindi contraddire l'ipotesi che gli stati cerebrali determinino gli stati mentali. Il punto di vista del "contenuto ristretto" afferma che gli stati mentali di un "mangiatore di hamburger" reale e di un "mangiatore di hamburger" di un cervello nella vasca sono i medesimi. Riflettendo relativamente alla possibilità del pensiero dell'intelligenza artificiale la posizione del "contenuto ristretto" appare appropriata dal momento che non ha senso affermare che la possibilità di pensiero dell'AI dipenda da condizioni esterne al sistema. Quello che conta è quindi lo stato cerebrale – ed il conseguente stato mentale ha una funzione precisa all'interno dell'attività mentale dell'entità pensante, sia questo un uomo o potenzialmente, una macchina.

Nella teoria funzionalistica quindi uno stato mentale è il risultato di un input determinato e due sistemi con le stesse logiche (il cervello umano ed un computer in grado di replicarlo perfettamente dal punto di vista neurale) sottoposti a medesimi input avrebbero medesimi stati mentali. Tale possibilità è approfondita dall'esperimento mentale di sostituzione del cervello. Immaginiamo che la tecnica e la tecnologia consentano di ricreare artificialmente e perfettamente i neuroni naturali e che fosse possibile sostituirli uno ad uno nella testa di una persona senza interrompere il funzionamento del cervello nel suo insieme. Quello che vorremmo osservare è sia il comportamento esterno del soggetto dell'esperimento, sia le sue modificazioni interne attraverso le sensazioni esperite dall'individuo. I difensori della teoria funzionalista sostengono che il soggetto dell'esperimento non avrebbe alcuna modificazione e che non vi sarebbe alcuna modifica neanche nella propria coscienza.

Un parallelismo interessante è quello della chimica organica e inorganica. Infatti prima del 1848 gli scienziati si interrogavano sulla possibilità di sintetizzare artificialmente una sostanza biologica e sulla possibilità che questa fosse indistinguibile da quella naturale; molti erano quelli che sostenevano che una tale ipotesi fosse irrealizzabile. Ma nel 1848 Frederik Wöhler sintetizzò artificialmente l'urea e questa era perfettamente identica al prodotto biologico smentendo con i fatti tutti coloro che ritenevano questo impossibile.

Nel caso dell'intelligenza artificiale non siamo ancora arrivati a tali risultati ma la logica potrebbe essere la medesima e quindi un comportamento apparentemente supportato da pensiero dovrebbe farci presupporre, così come

facciamo per gli uomini, che effettivamente all'apparenza corrisponda pensiero e coscienza.
Infine, le considerazioni di Turing suggeriscono che le differenze tra AI forte e AI debole si dissolveranno quando le macchine raggiungeranno un certo livello di sofisticazione.

3.6 Penrose

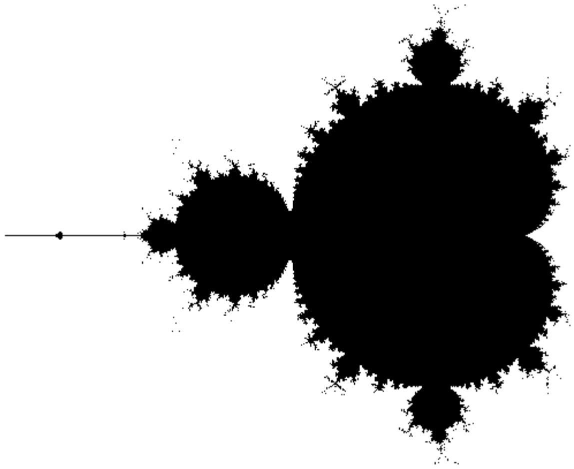
Sir Roger Penrose (Colchester, 8 agosto 1931) nega la possibilità dell'AI forte. Il matematico inglese vuole sottolineare come la struttura fisica, biologica e chimica del cervello umano abbia possibilità non replicabili da un computer digitale. Egli inoltre sostiene, pur condividendone la tesi, che gli esperimenti mentali di Searle, non siano stati in grado di spiegare perché oggetti biologici, come il cervello, hanno caratteristiche quali intenzionalità e semanticità che vengono negate a oggetti elettronici. La ricerca del matematico inglese si indirizza quindi nell'evidenziare «qualche nuovo principio che oggi manca del tutto»[21] ovvero un principio che giustifichi la differenza qualitativa tra la mente ed il calcolatore elettronico. In altre parole ci si chiede se i processi mentali siano di natura prettamente algoritmica, come sostiene chi crede nell'AI forte, o abbiano delle caratteristiche uniche rispetto ai calcolatori. Si ricorda che per "algoritmo" si intende un procedimento meccanico definito, un insieme di operazioni, volte alla risoluzione di un problema. Per una definizione di "meccanico" si rimanda alla descrizione della macchina di Turing e delle sue possibilità (e limiti).
Altro elemento fondamentale per la comprensione della tesi di Penrose, è la sua concezione della matematica, che lo stesso autore definisce platonica. Tale concezione deriva, tra le altre cose, da "scoperte" quali l'insieme di Mandelbrot[22].

---

[21] R. Penrose, *La mente nuova dell'imperatore*, trad. it. L. Sosio, Sansoni Editore, Roma 1997, pag. 39.

[22] Benoît Mandelbrot (Varsavia, 20 novembre 1924 – Cambridge, 14 ottobre 2010).

Infatti, il matematico franco-polacco, padre dei frattali, definì l'insieme che porta il suo nome da una semplice formula matematica. Da questa, attraverso l'elaborazione del computer, scaturisce la seguente rappresentazione grafica:



Penrose così commenta questa "scoperta": «Non importa quale computer venga usato per eseguire i calcoli (purché il computer funzioni regolarmente), a parte il fatto che differenze nella velocità e nella memoria del computer, e nelle sue capacità di visualizzazione grafica, possono condurre a differenze nella quantità di particolari fini che saranno rivelati e nella velocità con cui tali particolari saranno prodotti. Il computer viene usato essenzialmente nello stesso modo in cui il fisico sperimentale usa un'apparecchiatura per esplorare la struttura del mondo fisico. L'insieme di Mandelbrot non è un'invenzione della mente umana: esso fu una scoperta. Come il Monte Everest, l'insieme di Mandelbrot ha un'esistenza propria!»[23].

Vi è un ordine nel mondo (quello che lo stesso Mandelbrot cercava nella natura laddove gli altri vedevano solo caos), delle regole che esistono senza tempo e senza spazio. I matematici non sono quindi i creatori di mondi formali artificiali, invenzioni del pensiero umano, al contrario sono degli scopritori di strutture eterne ed indipendenti dall'uomo stesso: «nella matematica c'è qualcosa di

[23] R. Penrose, *La mente nuova dell'imperatore*, trad. it. L. Sosio, Sansoni Editore, Roma 1997, pag. 134.

assoluto e di *divino*»[24]. Tale concezione si oppone al formalismo che, di contro, sostiene che le strutture matematiche siano il prodotto del cervello umano caratterizzato da strutture e limiti propri. Il formalismo afferma come il ragionamento matematico debba necessariamente avere carattere algoritmico ovvero essere teoricamente "trascrivibile" in un insieme di regole attraverso lo svolgimento delle quali si può arrivare alle conclusioni possibili (anche se come abbiamo visto non tutto è dimostrabile). Queste regole possono, teoricamente, essere tradotte in algoritmi informatici.

Penrose prosegue l'esposizione della propria tesi mostrando esempi di matematica non "ricorsiva"[25], ovvero esempi di problemi nei quali le soluzioni (che umanamente siamo in grado di scovare) non possono derivare da un ragionamento meccanico di tipo algoritmico.

Ad esempio le soluzioni all'insieme di equazioni diofantee[26] non possono essere derivate da alcun algoritmo. Non esiste infatti alcun procedimento formalizzato in grado di trovare la soluzione $x = 13, y = 7, z = 2$ al seguente insieme di equazioni

$$z^3 - y - 1 = 0$$
$$yz^2 - 2x - 2 = 0$$
$$y^2 - 2xz + z + 12 = 0$$

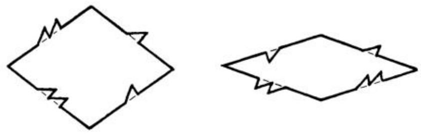
Un altro esempio è quello detto delle *tassellature*. La domanda, seppur in un ambito diverso, è sempre la stessa: esiste un procedimento per definire se sia possibile coprire un piano bidimensionale senza lasciare alcun spazio libero con le seguenti figure?

---

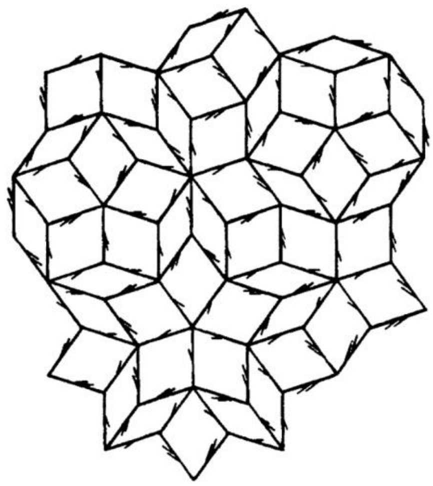
[24] R. Penrose, *La mente nuova dell'imperatore*, trad. it. L. Sosio, Sansoni Editore, Roma 1997, pag. 155

[25] Ricorsivo equivale in matematica ad "algoritmico" e "computabile" e denota operazioni matematiche che possono essere eseguite da macchine teoriche quali la macchina di Turing.

[26] dal nome del matematico greco del III secolo a.C. Diofanto.

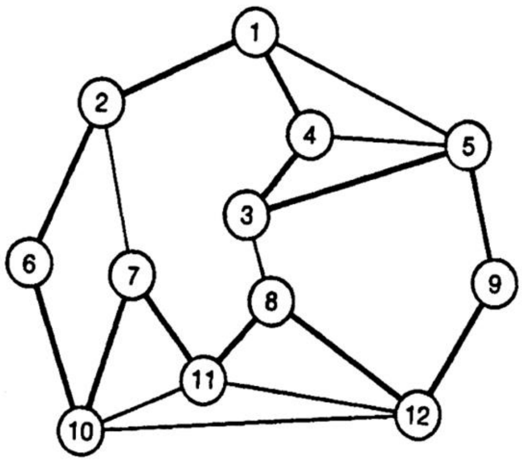


Anche in questo caso nel 1966 Robert Berger dimostrò che non esiste alcuna procedura di decisione del problema della tassellatura e che quindi anche questo ambito rientra nella matematica non ricorsiva. Non esiste quindi un algoritmo capace di trovare la soluzione a questo problema. Ne è invece capace la mente umana: a seguire la soluzione (anche se necessariamente parziale in quanto la domanda è posta per un piano infinito).



Ma vi è un'altra questione: anche per problemi in cui esista una soluzione algoritmica, è questa soluzione efficiente ed utilizzabile o tale procedura formalizzata ha tempi di elaborazione eccessivi? La teoria della complessità risponde appunto a questo interrogativo studiando l'evoluzione o meglio l'aumento del numero di passi degli algoritmi al crescere della "dimensione" del problema. Questi studi hanno quindi distinto sotto la lettera P i casi in cui se n è la dimensione del problema e N il numero di passaggi algoritmici necessario per risolverlo, all'aumentare di n, N cresce con ritmi che son al massimo multipli di $n, n^2, n^3, n^4, n^5, \ldots$

Se i problemi P hanno in qualche modo una complessità *trattabile* algoritmicamente, i problemi NP vedono una crescita di N al crescere di n tale da renderli di fatto inutilizzabili (qualora tali algoritmi fossero definiti). Vi sono esempi anche relativamente semplici di questi problemi: ad esempio la ricerca del *circuito di Hamilton* (ovvero un circuito chiuso formato solo da archi del grafo e che passa una e una sola volta per ogni vertice) all'interno di un grafo. A seguire l'esempio di un grafo dove le linee più spesse evidenziano il circuito di Hamilton.



In questo esempio all'aumentare di n, N cresce in modo tale da non rendere possibile l'elaborazione in tempi ragionevoli. Ad oggi non è stato trovato alcun algoritmo tale da poter classificare questo problema nella categoria P. Per problemi NP il tempo reale che occorrerebbe con grande valore di n cresce rapidamente fino a divenire maggiore dell'età dell'universo.

Problema analogo è quello classico del commesso viaggiatore; si vedrà in seguito la strategia adottata dalla moderna "intelligenza artificiale" per rispondere a questo quesito nel paragrafo 5.1.

Molti ritengono che sia impossibile per un qualsiasi dispositivo paragonabile a una macchina di Turing – un computer ad esempio – risolvere problemi NP e che, di fatto, NP e P siano "cose" di natura diversa, anche se poi, ad oggi, nessuno è stato in grado di esplicitare tale diversità. Insomma, la teoria della complessità aggiunge un problema a quello della computabilità: non tutto è ricorsivo e tra i problemi che sono algoritmicamente calcolabili, ve ne sono alcuni in cui, di fatto, tale logica procedurale è inutilizzabile[27].

Penrose affronta poi il tema della coscienza come parte essenziale dell'intelligenza e per individuare in essa quell' «ingrediente essenziale non algoritmico»[28]. Il fisico-matematico non si addentra in una precisa definizione di che cosa sia la coscienza "limitandosi" a constatare come questa sia una parte fondamentale della mente umana. L'intelligenza nello specifico si compone di una parte maggioritaria che correttamente può essere descritta algoritmicamente (e di conseguenza replicabile) e di una che procede con altri metodi. Vi è, inoltre, una gradazione nella coscienza, o consapevolezza, sia tra gli animali (è possibile assegnare una consapevolezza ad esempio agli scimpanzé o ai delfini, meno agli insetti), sia nella vita dei singoli, dall'infanzia alla vecchiaia, dal sonno alla veglia. La coscienza secondo l'autore è quella specifica capacità che ci permette di esprimere giudizi di verità, comprendere, apprezzare un'opera d'arte, condividere quello che solitamente viene definito senso comune. Essa non implica la precisa conoscenza delle ragioni che ci portano a formulare valutazioni come quelle appena menzionate. Ed infatti la possibilità di una precisa descrizione "computazionale" del perché di tali giudizi comporterebbe a ragione la possibilità teorica di una loro trascrizione algoritmica. Si potrebbe obiettare che vi sono meccanismi inconsci che regolano il nostro agire: l'opinione di Penrose è che, al contrario di quanto pensano i sostenitori dell'AI forte, tali automatismi inconsci, per quanto complicati ed oscuri, possano in linea di principio essere tradotti in un

[27] Si precisa che alcune nuove ipotesi sui computer quantistici lasciano aperta la possibilità della gestione in tempi ragionevoli di problemi ora classificati come NP. Si tratta ad oggi solo di un'apertura teorica a questa possibilità.

[28] R. Penrose, *La mente nuova dell'imperatore*, trad. it. L. Sosio, Sansoni Editore, Roma 1997, pag. 515.

programma. La differenza qualitativa tra computer e mente va ricercata nelle caratteristiche e nelle funzioni derivanti dalla coscienza (che come detto non implica la consapevolezza delle stesse). Ancora, la coscienza sarebbe quella «capacità di divinare (o distinguere *intuitivamente*), in circostanza appropriate, la verità dalla falsità (e la bellezza dalla bruttezza!)»[29]. Essa è quindi, per l'autore, quella speciale facoltà che entra in gioco nella nostra quotidianità ogni volta che formuliamo un giudizio. Ad esempio quando, senza esitazione, vengono sommate conoscenze, esperienze, impressioni sensoriali, decidiamo senza l'ausilio di procedure codificate.

Ma se al contrario ipotizzassimo che l'intero operare del cervello umano fosse algoritmico, potremmo chiederci come abbia fatto la natura a giungere a selezionare una specie caratterizzata da un così complicato sistema computazionale. In questa ipotesi la selezione naturale avrebbe quindi nel tempo - ed modo discontinuo - preferito individui con algoritmi più efficaci e quindi avvantaggiati nella sopravvivenza (e di conseguenza favorendo la trasmissione genetica del miglior "programma").

Anche concedendo tale possibilità, resta inspiegato il motivo per il quale la selezione naturale abbia dotato l'uomo di algoritmi che possano dare giudizi coscienti sulla *validità* di altri algoritmi. Tale ipotesi è per Penrose difficile da sostenere. Infatti in primo luogo è difficile ri-costruire l'algoritmo partendo dal suo output, e di conseguenza è improbabile che la selezione, che agisce necessariamente sui risultati, possa determinare l'evoluzione dell'algoritmo "sorgente". In secondo luogo, in algoritmi complessi, un piccolo mutamento in un ingranaggio potrebbe compromettere completamente l'efficacia del sistema. Tali conclusioni improbabili sono addotte da Penrose a sostegno della propria tesi nella convinzione che alla base dell'intelligenza vi siano delle idee fondanti e che

---

[29] R. Penrose, *La mente nuova dell'imperatore*, trad. it. L. Sosio, Sansoni Editore, Roma 1997, pag. 521.

queste possano esser colte e gestite solo attraverso la coscienza. Ancora, chiamando in causa il teorema di Gödel, i sostenitori dell'AI forte affermano che esista un algoritmo a noi oscuro dentro la nostra testa per provare la verità della proposizione di Gödel. Ma come è possibile che da un lato i matematici condividano un sistema formale che in maniera del tutto trasparente riesce a dar conto della matematica, mentre dobbiamo chiamare in causa un qualche cosa di misterioso per una proposizione che non possiamo dimostrare algoritmicamente? E com'è possibile che il logico austriaco sia stato in grado di scrivere tale proposizione se le oscure logiche di questa non possono essere svelate? Tale dimostrazione per assurdo, conclude l'autore, conferma il fatto che vi siano verità matematiche non algoritmiche.

Altre argomentazioni a sostegno della tesi del procedere non computazionale della mente vengono dalle osservazioni (anche in campi diversi del sapere quali le arti) di esperienze quali l'ispirazione e l'intuito. In questi casi sembra quasi che una visione si affacci nella mente del fortunato soggetto che in un attimo riesce ad afferrare quello che, magari, da tempo cercava. Oppure in un attimo coglie la "bellezza" di una visione matematica così come di un componimento musicale la cui visione è chiara nella mente ben prima che questa venga trascritta su di uno spartito.

Penrose prosegue osservando che la comunicazione dei concetti matematici tra i matematici stessi ha caratteri particolari. Essa infatti va oltre le parole e gli specifici approcci dei singoli e si basa sul comune "contatto" con il mondo platonico delle idee. Sarebbe infatti impossibile una comunicazione così efficace tra persone diverse per linguaggi e approcci se non presupponendo un mondo comune e l'accesso a questo da parte dei singoli pensatori (il cui intendersi poggia sulla medesima visione dell'idea eterna). La "bellezza", sostantivo con il quale a volte gli scienziati definiscono una scoperta o una formula, farebbe quindi riferimento all'armonia insita nel mondo (delle idee) e alla contemplazione quando si riesce a coglierla.

Ancora, è per il matematico inglese quasi impossibile credere che teorie quali la relatività o la fisica quantistica siano il frutto elaborato di un algoritmo complessissimo che le ha create; al contrario, nella sua concezione platonica dell'esistenza delle realtà matematiche e fisiche, queste sono state svelate all'umanità da alcune menti particolarmente capaci.

Per concludere, la scoperta relativamente recente dei quasi-cristalli offre una possibile chiave di volta per una risposta alle domande su coscienza e intelligenza. I quasi-cristalli, infatti, hanno una disposizione cristallina particolare composta da tassellature (si ricordi quanto detto sopra) con una composizione non locale. Per combinare così questi cristalli è necessario esaminare lo stato della tassellatura a molti atomi di distanza per non incorrere in errori e "spazi vuoti" nella composizione dei tasselli. Penrose propone delle ipotesi di fisica quantistica per spiegare il perché della crescita ordinata di strutture che non possono ingrandirsi semplicemente per accrescimento, ma la cui evoluzione deve avere una qualche forma di governo, e ipotizza che queste stesse logiche siano presenti nel cervello. Questo infatti, a differenza del computer, è una struttura in continua evoluzione attraverso l'attivazione e la disattivazione delle sinapsi, la crescita o contrazione delle spine detritiche, dei collegamenti tra neuroni. Similmente a quanto avviene nei quasi-cristalli, l'ignota teoria quantistica GQC[30] (che non viene qui approfondita) rappresenterebbe quindi una soluzione alla gestione delle complessità cerebrali fino a poter offrire una spiegazione scientifica a fenomeni complessissimi quali desideri ed intenzioni. Inoltre, essendo il problema della tassellatura non risolvibile algoritmicamente, sarebbe una conferma della struttura non computazionale del cervello e della conseguente impossibilità dell'AI forte.

---

[30] La GQC, acronimo di *gravità quantistica corretta* è un'ipotesi teorica avanzata da Penrose per dar risposta ai molti quesiti aperti della fisica quantistica.

Infine, per il matematico inglese la coscienza è «la *visione* di una verità [...] che possa rappresentare una qualche sorta di reale contatto col mondo platonico dei concetti matematici ideali»[31], ed una visione di queste realtà (le idee) atemporali ed eterne. Tale ipotesi potrebbe quindi ben spiegare esperienze quali l'intuito o l'illuminazione sperimentate da matematici o artisti. Ne è un esempio la testimonianza di Mozart che raccontava la "creazione" dei propri spartiti come una visione unitaria dell'insieme; non quindi una creazione, ma appunto una scoperta. Ma, conclude Penrose, la coscienza non è spettatrice passiva e semplice collegamento tra il mondo esterno ed il mondo atemporale delle idee; essa svolge piuttosto un ruolo attivo e fondamentale connesso ad un forte vantaggio selettivo. Anche in questo caso l'autore non può che formulare l'ipotesi che la GQC consenta un'azione attiva della coscienza.

Resta "la sorpresa e l'ammirazione" per il cervello e la sua organizzazione e capacità. A questa si accompagna il riconoscimento del poco che ancora si sa sulla coscienza e l'apertura (e il timore) per il futuro: «Se mai riusciremo a scoprire nei particolari quale sia quella particolarità che permette ad un oggetto fisico di diventare cosciente, potremmo essere concepibilmente in grado di costruire noi stessi degli oggetti coscienti, anche se essi non potrebbero certo essere chiamati "macchine" nel senso in cui noi oggi intendiamo questa parola. [...] Si potrebbe anche immaginare che [...] tali oggetti[32] potrebbero riuscire a soppiantare realmente gli esseri umani, mentre (nell'opinione di coloro che la pensano come me) i computer algoritmici sono condannati ad operare al nostro servizio»[33].

---

[31] R. Penrose, *La mente nuova dell'imperatore*, trad. it. L. Sosio, Sansoni Editore, Roma 1997, pagg. 562-563.

[32] Tali aggetti coscienti sarebbero inoltre "liberi" dall'eredità storica umana e quindi, in linea di principio, senza istinti e condizionamenti legati agli aspetti evolutivi, di conservazione della specie e culturali.

[33] R. Penrose, *La mente nuova dell'imperatore*, trad. it. L. Sosio, Sansoni Editore, Roma 1997, pag. 525.

4 Libertà e determinismo

Una delle cose che sembrano distinguere gli uomini dalle macchine è la presenza nei primi della libertà. Il libero arbitrio è la possibilità di svincolarsi dalla rigida ed invincibile catena causa-effetto ed il potere di decidere "senza causa".

La fisica classica lega con il vincolo necessario delle leggi fisiche ogni accadimento nel mondo. La fisica quantistica introduce una componente casuale. Come osserva lo stesso Penrose, tale casualità (ed anche gli elementi non algoritmici dallo stesso proposti) non migliora di molto la possibilità del libero arbitrio e tutt'al più «...implicherebbe che il futuro non sia computabile a partire dal presente, anche se potrebbe esserne determinato»[34].

Simili sono le riflessioni di Einstein nella lettera indirizzata a Ernst Straus: «Quel che mi interessa veramente è se Dio avrebbe potuto fare il mondo in un modo diverso; ossia se la necessità della semplicità logica può lasciare qualche libertà!».

Ai fini del presente lavoro, e relativamente al problema del libero arbitrio, sono interessanti le riflessioni di Spinoza, la formalizzazione delle stesse di Boole e la proposta di definizione di Chaitin circa il concetto di casualità. Viene infine presentata la soluzione pragmaticamente adottata in informatica per la generazione di sequenza di numeri "casuali" visto che, ad oggi, non si può chiedere al computer di darci dei numeri "a caso".

Resta la domanda se l'uomo sia in grado di dare numeri "a caso".

4.1 Spinoza

Nucleo principale del pensiero di Baruch Spinoza (Amsterdam, 24 novembre 1632 – L'Aia, 21 febbraio 1677) è la definizione di sostanza, così definita: «Per sostanza intendo ciò che è in sé e che per sé si concepisce: ossia ciò il cui concetto non ha

---

[34] Ivi, pag. 545.

bisogno di un'altra cosa da cui si debba formare»[35]. Questo concetto non ha nulla di nuovo e si rifà ad una lunga tradizione di origine aristotelica, largamente condivisa e accettata.

La sostanza è quindi caratterizzata da assoluta indipendenza ontologica. Ne consegue che ciò che è da altro dipendente non può essere definito tale. Nel pensiero cartesiano erano state definite tre sostanze: *res cogitans*, *res extensa* e Dio. Le prime, anche se reciprocamente indipendenti, sono dipendenti da Dio quale loro creatore e poco conta il fatto che dopo la creazione queste possano dirsi libere dal creatore. Restano, infatti, ontologicamente dipendenti da Dio e quindi non possono essere dette sostanze. Il rigore spinge l'autore a fondare la metafisica "prima" del *cogito* e del mondo, indagando ciò da cui questi dipendono, appunto, la vera sostanza.

Inoltre, la sostanza è infinita in quanto dotata di infiniti attributi, ognuno dei quali è infinito nel suo genere. Il contrario implicherebbe che «dovrebbe essere delimitata da un'altra della medesima natura, che a sua volta dovrebbe necessariamente esistere; ci sarebbero quindi due sostanze del medesimo attributo, il che è assurdo» (per la Proposizione V)[36].

Riassumendo la sostanza deve essere:

- indipendente ontologicamente;
- infinita;
- perfetta.

Ci si potrebbe ora chiedere se questa sostanza esiste o se potrebbe non esistere. A questo dubbio risponde la proposizione VII che afferma «Alla natura della sostanza appartiene l'esistere»[37] e ciò è confermato dalla relativa dimostrazione «una sostanza non può essere prodotta da altro; sarà dunque causa di sé, cioè la

[35] B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014, p. 27, I,d3.

[36] Ivi, p. 31, I,p8,dim.

[37] B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014, p. 31, I,p7.

sua essenza implica necessariamente l'esistenza, ovvero alla sua natura appartiene l'esistere»[38].

Nelle prime pagine dell'opera l'autore parla genericamente delle caratteristiche delle sostanze. Ma una volta che queste sono state definite, non possiamo che concludere che può esistere solo una sostanza in quando essendo questa infinità non può essere limitata da altro.

«Dunque esiste una sola sostanza assolutamente infinita, e poiché tutti sono d'accordo nell'identificare Dio con ciò che è assolutamente infinito, questa sostanza è Dio»[39].

Da definizioni iniziali sulle quali non v'è motivo di dubitare non possiamo che concludere che l'unica vera cosa che esiste è una sola sostanza, che questa si identifica con Dio, eterno, perfetto, infinito e che tutto è in Dio. E la proposizione XIV non può che confermare che «Oltre a Dio non può esserci né essere concepita alcuna sostanza»[40].

Prosegue Spinoza: Proposizione XV «Tutto ciò che è, è in Dio, e nulla può essere né essere concepito senza Dio»[41]. Il mondo, il pensiero, l'universo esistono solo in Dio.

Precedentemente è stato introdotto il concetto di attributo senza però darne una definizione precisa: «Definizione IV – Per attributo intendo ciò che l'intelletto percepisce della sostanza in quanto costitutivo della sostanza»[42]. Gli attributi sono i predicati essenziali di Dio-sostanza (la sostanza non sarebbe tale senza di essi). Questa è quindi compresa dall'intelletto attraverso i modi di darsi della sostanza, ovvero, gli attributi. Si precisa poi, però, che l'intelletto finito dell'uomo può concepire solo due degli infiniti attributi: ovvero pensiero ed estensione. La

---

[38] Ivi, p. 31, I,p7,dim.

[39] E. Scribano, *Guida alla lettura dell'Etica di Spinoza*, Editori Laterza, Bari 2008.

[40] B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014, p. 41, I,p14.

[41] Ivi, p. 41, I.p15.

[42] Ivi, p. 27, I,d4.

distanza da Cartesio, e da tutta la tradizione classica, non potrebbe essere maggiore: pensiero ed estensione non esistono di per sé, non sono sostanze ma solo attributi di Dio, o meglio sono i soli due attributi, tra gli infiniti, che l'umano intelletto può percepire della sostanza.

In altre parole, se l'uomo si rivolge verso la sostanza dal punto di vista dell'estensione questa sarà vista come *res extensa*, mentre dal punto di vista del pensiero sarà sostanza pensante. Non si tratta quindi di due entità distinte ma della medesima e ciò che cambia sono solo gli attributi attraverso i quali guardiamo alla sostanza.

Quindi Dio è anche estensione. Qui è il grande scandalo del pensiero di Spinoza. Dio non è altro dal mondo: il mondo, la materia sono Dio (una parte). Dio è la totalità delle cose (come detto anche pensiero ed estensione) e tale conclusione è necessaria in quanto a queste conclusioni siamo giunti in forza del ragionamento logico in un processo argomentativo "geometrico".

Va maggiormente chiarito il rapporto tra pensiero ed estensione. Come abbiamo detto questi sono due attributi, due modi di vedere la medesima ed unica sostanza. Essi quindi si riferiscono, o meglio sono, la stessa cosa. Il dualismo che opponeva pensiero ed estensione è tramontato: «Proposizione VII – L'ordine e la connessione delle idee è lo stesso che l'ordine e la connessione delle cose»[43], e più avanti nello scolio successivo «…la sostanza pensante e la sostanza estesa sono un'unica e medesima sostanza, che si comprende ora sotto questo, ora sotto quell'attributo»[44].

Ancora, non c'è comunicazione tra i due attributi ovvero il pensiero non può determinare l'estensione e simmetricamente l'estensione non ha alcuna influenza sul pensiero. È qui utile un esempio: quando voglio muovere la mano, ed in effetti vedo la mano muoversi, non è il pensiero la causa del movimento ma le due cose

[43] Ivi, p. 87, II,p7.

[44] Ivi, p. 87, II,p7,s.

(il movimento della mano nello spazio e la volontà di muoverla) sono parallele, simultanee ed in-comunicanti. Nel medesimo tempo vedo la mano muoversi dal punto di vista della sostanza attraverso l'attributo dell'estensione e percepisco la volontà di muovere la mano, con gli "occhi" dell'attributo del pensiero: «Proposizione II – Né il corpo può determinare la mente a pensare, né la mente può determinare il corpo al moto o alla quiete né ad alcunché d'altro (posto che ci sia)»[45].

Che il pensiero non possa avere alcun effetto sulla materia implica la necessità di tutte le cose e sia la natura che la nostra mente seguono un ordine necessario di eventi; non si da alcuna libertà e «… e dunque queste decisioni della mente si originano nella mente con la stessa necessità delle idee delle cose esistenti in atto. Quelli dunque che credono di parlare, o di tacere, o di fare qualsiasi cosa per libera decisione della mente, sognano ad occhi aperti»[46].

Tutto è governato da un rigido determinismo. Tutto ha una causa e da questa non possono che discendere effetti necessari. Non si dà alcuno spazio al libero arbitrio umano. L'unica cosa libera è la sostanza ma anche in questo caso si tratta di un'accezione di libertà molto diversa da quella abitualmente in uso che rimanda alla contingenza della cose: «Definizione VII – Sarà detta libera quella cosa che esiste per la sola necessità della sua natura e che da sola si determina ad agire; necessaria invece, o meglio costretta, quella che è determinata da altro a esistere e ad operare per una certa e determinata ragione»[47]. Solo Dio può dirsi libero: «Ne deriva in secondo luogo che Dio solo è causa libera. Dio solo, infatti, esiste in base alla sola necessità della sua natura, e in base alla sola necessità della sua natura agisce. Dunque è da solo causa libera»[48].

---

[45] Ivi, p. 155, III,p2.

[46] Ivi, p. 161, III,p2,s.

[47] Ivi, p. 27, I,d7.

[48] Ivi, p. 49, I,d17,c2.

La sostanza è quindi l'unica cosa libera ma, come anticipato, la libertà non va intesa come arbitrarietà o discrezionalità. Dio è libero nel senso che non è determinato da altro se non da sé: il dispiegarsi della sostanza (anche nel pensiero e nell'estensione) non è causato da altro all'infuori della sostanza stessa. Ma questo "svolgersi" della sostanza segue un ordine necessario. Inoltre, questa necessità non etero-determinata della sostanza è coerente con le caratteristiche di perfezione e onnipotenza di Dio. Se così non fosse potremmo giungere alla conclusione che Dio non può fare tutte le cose che potenzialmente potrebbe fare (infatti ogni azione in una vera scelta implicherebbe l'impossibilità dell'alternativa non perseguita) e che quindi Dio «... non possa fare tutte le cose alle quali la sua potenza si estende; e non vedo che cosa ci si possa figurare di più assurdo e di più contrastante con l'onnipotenza di Dio»[49]. La libera necessità della sostanza è esemplificata dalla natura del triangolo nel quale, da tutta e per tutta l'eternità, la somma degli angoli interni non potrà che essere pari ad un angolo piatto. La sostanza è libera ma in essa tutto è eternamente determinato e necessario.

Con la sua concezione Spinoza stravolge la concezione tradizionale di Dio: Egli non è più un elemento esterno al mondo, non ha creato il mondo, e non osserva e giudica gli uomini: Egli è la verità di tutte le cose e non è fuori dal mondo, anzi è il mondo spogliato di qualsiasi elemento antropologico.

Nel pensiero del filosofo olandese, i corpi estesi sono particolari modalità dell'attributo dell'estensione e le idee e gli intelletti finiti sono modalità dell'attributo del pensiero. Ancora, tutto ciò che accade, in quanto finito è necessitato dalla sostanza e non ha alcuna autonomia: Proposizione XXIX «Nella natura delle cose non si dà nulla di contingente, ma tutte le cose sono determinate dalla necessità della natura divina a esistere e ad operare in un certo modo»[50].

---

[49] Ivi, p. 49, I,p7,s.

[50] Ivi, p. 61, I,d29.

In Spinoza scompare anche il concetto di creazione. Tradizionalmente questa era un atto libero che rendeva esistente qualcosa di altro (il mondo - l'uomo) dal Dio stesso. Nel nostro non c'è spazio per una simile concezione: il mondo non può essere creato perché è già eternamente e necessariamente in Dio. Ne segue la proposizione XVIII: «Dio è causa immanente, non transitiva, di tutte le cose»[51]. Se infatti la causa transitiva scinde la causa dall'effetto (l'artigiano crea il proprio manufatto che è altro da sé), nella causa immanente, invece, causa ed effetto coincidono e in Dio la causa rimane identica ai suoi effetti e di fatto coincidono nell'unica sostanza possibile (*causa sui*).

Va sottolineato che il mondo è totalmente dipendente da Dio; il mondo non ha in sé le ragioni della sua esistenza. All'accusa di ateismo che spesso viene rivolta a Spinoza si può quindi opporre un punto di vista contrario ovvero un atteggiamento teistico: ogni cosa è Dio e il mondo stesso in Egli si dissolve.

Il Dio di Spinoza è l'insieme di tutte le cose, da sempre e per sempre necessarie.

Ma perché l'uomo ha sempre fallito nell'identificare correttamente la natura di tutte le cose e questo ragionamento logico è sempre sfuggito a saggi e filosofi? L'autore ritiene che la causa di tutti gli errori sia da rintracciarsi nei pregiudizi.

«La radice di tutti i pregiudizi risiede in tre caratteri antropologici specifici: gli uomini nascono ignari delle cause delle cose, ricercano il proprio utile, ne sono consapevoli. Da queste premesse seguono necessariamente due pregiudizi fondamentali dai quali derivano tutti gli altri: l'illusione di essere liberti e il ragionamento finalista»[52].

Causa del primo pregiudizio, ovvero dell'illusione della libertà, è il fatto che l'uomo è perfettamente consapevole dei propri desideri e della propria volontà. Ciò di cui non è consapevole solo le cause delle proprie volizioni e l'uomo si guarda ben dall'indagarle. Ne consegue che ritiene essere libero nella propria

---

[51] Ivi, p. 51, I,d18.

[52] Ivi, p. 71, I,app.

volontà in quanto solo di questa è consapevole, accettando quindi l'assurdo che non vi siano cause che la determinano. Scambia quindi gli effetti per le cause e così pensando ritiene, sbagliando, di essere libero di scegliere quando in realtà è determinato dai desideri e dalle passioni. La libertà è solo l'illusione creata dall'ignoranza.

Relativamente al secondo grande pregiudizio Spinoza spiega: «… gli uomini fanno ogni cosa in vista di un fine, ovvero dell'utile a cui mirano; quindi cercano sempre di conoscere le cause finali di ciò che è avvenuto, e non appena se le sentono dire si tranquillizzano, ovviamente perché non hanno alcun ulteriore motivo di dubbio. Se invece non riescono a sentirle da altri, non resta loro che ripiegarsi su sé stessi, e riflettere sui fini per i quali essi stessi sono soliti determinarsi a cose del genere, e così, di necessità, giudicano del carattere degli altri in base al proprio»[53].

Con l'introduzione di riti, culti e istituzioni poi questi errori si erigono a vere superstizioni delegando ad un'entità superiore le cose che sfuggono all'umana comprensione e rendendo i pregiudizi ancora più difficili da scalfire.

4.2 Spinoza e la logica di Boole

L'*Ethica ordine geometrico demostrata* è il primo esempio dell'utilizzo del metodo logico matematico applicato ad un lavoro filosofico. L'opera del filosofo olandese si presta a questo tipo di analisi dal momento che consta di definizioni, assiomi e dimostrazioni, anche se Boole lamenta la presenza, nell'opera di Spinoza, di definizioni vaghe e poca chiarezza. Procediamo ora all'analisi logica dell'opera.

Assioma I: *Omnia, quae sunt, vel in se, vel in alio sunt*[54]

Questo può essere tradotto nella coppia di proposizioni:

---

[53] L. Vinciguerra, *Spinoza*, Carocci Editore, Roma 2016, pag. 51.

[54] «Tutte le cose che sono o sono in sé o sono in altro» B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014, p. 29.

$$x = cose\ che\ sono\ in\ sé$$

$$x' = cose\ che\ sono\ in\ altro$$

Queste, visto che 1 rappresenta l'universo possono essere scritte come

$x + x' = 1$ che rappresenta appunto la traduzione simbolica dell'assioma I.

L'equazione può essere anche scritta come

$$x = 1 - x'$$

Assioma II: *Id quod per aliud non potest concipi, per se concipi debet*[55]

Con la stessa logica possiamo scrivere:

$$y = cose\ che\ possono\ essere\ concepite\ in\ sé$$

e

$$y' = cose\ che\ possono\ essere\ concepite\ in\ altro$$

L'assioma II è rappresentato dall'equazione

$$y = 1 - y'$$

Proseguendo: la definizione III afferma che per sostanza si intente "ciò che è in sé e per sé si concepisce: ossia ciò che il cui concetto non ha bisogni del concetto di un'altra cosa da cui si debba formare" e la definizione V chiarisce il termine modo ovvero "le affezioni della sostanza, ossia ciò che è in altro, tramite il quale anche si concepisce". Se chiamiamo z la sostanza e z' i modi, ne consegue

$$z + z' = 1 \text{ o } z = 1 - z'$$

che può esser letta come "le sostanze sono tutte le cose che non sono modi".

Ancora, dalla definizione VII[56] e chiamando f le cose libere e f'le cose necessarie,

$$f = 1 - f'$$

---

[55] «Ciò che non può essere concepito per altro, deve essere concepito per sé» B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014, p. 29.

[56] «Sarà detta libera quella cosa che esiste per la sola necessità della sua natura e che da sé sola si determina ad agire; necessaria invece, o meglio costretta, quella che è determinata da altra a esistere e ad operare per una certa e determinata ragione» B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014, p. 27.

La definizione I[57] e l'assioma VII[58] possono essere sintetizzati dicendo che l'universo si compone di cose che sono cause e sono esistenti da sé (e) e cose che causate da altro (e'). Similmente a quanto visto sopra questa può essere così formulata:

$$e = 1 - e'$$

La definizione III stabilisce l'identità tra la sostanza e ciò che si concepisce per sé ovvero con la medesima simbologia

$$z = y$$

L'assioma IV[59] stabilisce l'identità tra causa e con ciò per cui la cosa è concepita:

$$y = e$$

La definizione VII afferma l'identità tra una cosa libera e una cosa la cui esistenza dipende da sé:

$$f = e$$

La definizione V[60] identifica il modo z' con ciò che è in altro da sé (x')

$$z' = x'$$

Se consideriamo l'equazione dell'assioma I $x = 1 - x'$ e quest'ultimo risultato e sostituiamo x' con z' otteniamo

$$x = 1 - z'.$$

Se affianchiamo questa alla precedente

$$z = 1 - z'$$

---

[57] «Per causa in sé intendo ciò la cui essenza implica l'esistenza; ovvero ciò la cui natura non può esser concepita se non come esistente» B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014, p. 27.

[58] «Se qualcosa può essere concepito come non esistente, la sua essenza non implica l'esistenza» B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014, p. 29.

[59] «La conoscenza dell'effetto dipende dalla conoscenza della causa e la implica » B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014, p. 27.

[60] «Per modo intendo le affezioni della sostanza, ossia ciò che è in altro, tramite il quale anche si concepisce » B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014, p. 27.

possiamo concludere che

$$z = x$$

Ora, combinando insieme tutti i risultati abbiamo:

$$x = y = z = f = e = 1 - x' = 1 - y' = 1 - f' = 1 - z' = 1 - e'$$

da questa ricaviamo

$$z = 1 - e'$$

che rappresenta la proposizione VI: *Una substantia non potest produci ab alia substantia*[61]. Analogamente

$$z = e$$

è la traduzione simbolica della proposizione VII: «Alla natura della sostanza appartiene l'esistere»[62].

L'analisi logica di Boole procede poi evidenziando quelli che, secondo l'autore, sono degli errori del ragionamento del filosofo olandese dipendenti *"soprattutto dall'uso ambiguo delle parole"*. Nello specifico la critica è rivolta all'utilizzo di "finito" e "finito nel suo genere". Nella Definizione II infatti «si dice finita nel suo genere quella cosa che può essere delimitata da un'altra della medesima natura»[63]. E poiché «nella natura delle cose non ci possono essere due o più sostanze della stessa natura o attributo»[64] (Proposizione V), «Alla natura della sostanza appartiene l'esistere»[65] (proposizione VII), ne consegue che esiste una sola sostanza infinita. Boole critica il concetto sottointeso nella dimostrazione, ovvero che l'idea di spazio che rappresenti l'unica forma della sostanza e tutte le cose esistenti siano affezioni dello spazio.

---

[61] «Una sostanza non può essere prodotta da un'altra sostanza» B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014, p. 31.

[62] B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014, p. 31.

[63] Ivi,, p. 27.

[64] Ivi, p. 29.

[65] Ivi, p. 31.

Per l'autore questa concezione, fondamento della costruzione spinoziana, è debole in quanto si potrebbero concepire sostanze senza l'attributo dello spazio e quindi evitando la confusione tra "finito" e "finito nel suo genere".
L'analisi del logico quindi se è da una parte riesce a tradurre la trattazione filosofica in strumenti formali e rigorosi, dall'altra ne evidenzia i limiti (del pensiero di Spinoza). La ragione di questi è, per Boole, connessa all'impossibilità di comprendere l'infinito partendo dal finito. I limiti dell'intelletto e della conoscenza rendono inesorabilmente vani i tentativi di raggiungere la certezza dell'esistenza di un Essere infinito; non resta, secondo l'autore, che accontentarsi di raccogliere indizi dal mondo per giungere alla conclusione che è probabile esista un reggitore del mondo e causa intelligente dello stesso.

4.3 Chaitin

Gregory John Chaitin (Chicago, 25 giugno 1947) è un matematico ed informatico. L'unione di questi ambiti di ricerca ha generato considerazioni interessanti fondendo la matematica pura con un approccio più pragmatico tipico di chi ha a che fare con la programmazione. Inoltre traducendo l'idea di Leibniz che «Dio ha scelto il [mondo] più perfetto, cioè quello che è a un tempo il più semplice quanto a ipotesi e il più ricco di fenomeni»[66] nella logica digitale, Chaitin sostiene che «una teoria scientifica va pensata come un programma, scritto in binario, per calcolare le osservazioni, anch'esse scritte in binario. Ed esiste una legge di natura se vi è compressione, se i dati sperimentali sono compressi in un programma che ha meno bit dei dati che spiega. Maggiore è il grado di compressione, migliore è la legge, meglio si capiscono i dati»[67].

---

[66] G. Leibniz, *Philosophical Essays*, R. Ariew e D. Garber, Hachett, Indianapolis, 1980, pag. 69.

[67] G. Chaitin, *Alla ricerca di omega*, trad. it. S. Frediani, Adelphi, Milano 2007, pag. 55.

Ed il programma di Hilbert, tradotto in termini informatici, equivaleva alla dimostrazione dell'esistenza di un unico programma (il sistema assiomatico formale) capace, in matematica, di spiegare tutte le cose senza alcuna contraddizione.

Le dimostrazioni di Gödel e Turing confermarono che il sogno di Hilbert era impossibile e la matematica incompleta.

L'approccio informatico alla matematica conduce anche a considerazioni originali per quanto riguarda il concetto di casualità.

La riflessione di Chaitin su questo tema parte dalla dimostrazione dell'impossibilità di una definizione di casualità di Borel[68]: ipotizziamo di riuscire in qualche modo a distinguere una serie di numeri i cui decimali formano una successione casuale da altri numeri che non hanno questa caratteristica. Prendiamo ora il primo numero di N numeri che soddisfano le specifiche della *casualità*. Ora questo numero avrebbe una caratteristica particolare in quanto sarebbe il primo numero tra i numeri N con questa proprietà. La casualità sarebbe quindi una caratteristica comune dei numeri N tale da permetterci di distinguerli dagli altri. Tale ragionamento potrebbe proseguire ipotizzando "gradi" di casualità (si potrebbe quindi proseguire selezionando un sottoinsieme di numeri con una ulteriore specifica casualità, e così via). Ma tale conclusione è impossibile: casuale significa «indistinguibile dalla folla, privo di caratteristiche peculiari»[69]. Una definizione, per questa via, è quindi impossibile. All'avvicinarsi della definizione di casualità questa sfugge e non potrebbe essere diversamente visto che essa rappresenta un fenomeno privo di regole. Come distinguere, quindi, ciò che ha regole da ciò che non ne ha, visto che ciò implicherebbe la dimostrazione

---

[68] Félix Edouard Justin Émile Borel, matematico francese (Saint-Affrique, 7 gennaio 1871 – Parigi, 3 febbraio 1956).

[69] G. Chaitin, *Alla ricerca di omega*, trad. it. S. Frediani, Adelphi, Milano 2007, pag. 101.

della non esistenza di regole? E d'altra parte come si può esser certi che la sequenza di decimali di un numero che pare casuale non sia invece il risultato di una regola che ancora non è stata scoperta?
Il matematico ed informatico americano verificò anche che è impossibile dimostrare che un certo programma è il più breve (richiamandosi alla teoria di Leibniz) per l'ottenimento di un dato risultato (è interessante notare che egli non utilizza la parola breve ma "elegante" quasi a voler rafforzare l'ideale estetico della sintesi matematica).
Da queste premesse, Chaitin propone la propria idea di casualità come incomprimibilità algoritmica. Se quindi la dimensione in bit del miglior programma trovato per un dato risultato è superiore alla dimensione in bit del risultato stesso, allora possiamo concludere che ciò che cerchiamo di spiegare è inspiegabile ovvero casuale, ovvero incomprimibile algoritmicamente. Si nota come questa definizione sia più restrittiva rispetto a quella che indica come casuale un numero la cui serie decimale sembra statisticamente indipendente e uniforme. Ne è un esempio π che è calcolabile e quindi comprimibile.

4.4 Successioni pseudo-casuali

Spesso nei sistemi informatici è necessario disporre di successioni pseudo-casuali. Gli utilizzi di tali successioni variano molto dalla crittografia agli studi statistici e probabilistici su determinati eventi, alle simulazioni di modelli matematici, macroeconomici o, ad esempio, delle previsioni del tempo.
La soluzione individuata in informatica è detta PRNG (acronimo dell'inglese pseudo-random number generator). Si tratta di un algoritmo deterministico in grado di prendere una sequenza random in ingresso di lunghezza k e restituirne in uscita una di lunghezza $l > k$ che sembri casuale. Una sequenza di numeri per esser definita casuale deve possedere due caratteristiche primarie: distribuzione

uniforme ed indipendenza[70]. Nel caso delle sequenze binarie, ovvero costituite da 0 e 1, questo si traduce nel fatto che la quantità di 0 non differisca molto dalla quantità di 1 e che non sia possibile trovare una qualche relazione che ci permetta di legare i numeri adiacenti alla stringa tramite una relazione.

Uno dei primi metodi proposti per generare una serie pseudo-casuale fu quello suggerito dal John von Neumann (Budapest, Ungheria 28 dicembre 1903 - Washington, Stati Uniti 8 febbraio 1957) nel 1946. L'idea del matematico era estremamente semplice: si prende un numero iniziale qualunque di 10 cifre, questo viene elevato al quadrato e si selezionano le 10 cifre centrali di questo nuovo numero. Queste sono il nuovo numero pseudocasuale e tale processo può essere ripetuto fino alla generazione della lunghezza della stringa voluta. Ovviamente tale procedura può essere utilizzata anche per numeri con un numero di cifre diverso. Questo metodo è sicuramente veloce ma ha dei limiti: ad esempio se durante l'iterazione del processo i numeri mediani sono 0, tutti i successivi numeri pseudo-casuali saranno composti di zero.

Una sottoclasse dei PRNG sono i *Cryptographically Secure Pseudo-random Number Generator* che si caratterizzano per una maggior sicurezza[71] e quindi utili ad utilizzi criptografici.

Nel tempo si sono sviluppati algoritmi molto complessi ed efficaci in grado di generare sequenze affidabili e sicure, sempre utilizzando un dato iniziale detto seme, ed elaborando lo stesso attraverso formule calcoli deterministici. Una seconda via per riproporre una selezione di numeri pseudo-casuali è quella di "far lanciare i dadi al computer". Tale approccio è detto TRNG (dall'inglese *true*

---

[70] Per uniforme s'intende che la distribuzione ha la medesima probabilità su tutti i punti dell'intervallo; per indipendente s'intente un evento non influisce su un altro evento e nel nostro caso il verificarsi di un numero non ha alcun effetto sui numeri possibili successivi.

[71] Per maggior sicurezza si intende la maggior difficoltà nel ricostruire l'algoritmo partendo dalla successione generata dallo stesso algoritmo.

random number generator). La strategia di questi generatori di numeri pseudo-casuali è quello di estrarre la casualità da fenomeni fisici.
Tali dispositivi sono basati tipicamente su fenomeni microscopici come il rumore termico o l'effetto fotoelettrico o altri fenomeni quantistici. Si tratta di processi stocastici che sono, in teoria, completamente imprevedibili.
Un generatore hardware TRNG solitamente consiste in un trasduttore in grado di convertire alcuni aspetti di un fenomeno fisico stocastico in segnale elettrico, un amplificatore e altri circuiti elettrici necessari ad incrementare l'ampiezza delle fluttuazioni casuali rilevate ad un livello misurabile e un convertitore analogico-digitale per tradurre l'output in un segnale digitale (spesso un semplice segnale binario di 0 e 1). Attraverso il campionamento ripetuto delle variazioni del segnale si ottiene una serie pseudo-casuale di numeri.
I generatori di numeri pseudo-casuali possono anche essere ottenuti da fenomeni macroscopici, usando strumenti come carte da gioco, dadi, la ruota della roulette e le macchine da lotteria. La impredicibilità è spiegata dalla teoria dei sistemi dinamici instabili e dalla teoria del caos. Anche se i fenomeni macroscopici sono deterministici nell'ambito della meccanica newtoniana, i risultati di uno strumento ben progettato, come la ruota della roulette, non possono essere in pratica predetti poiché dipendono in modo sensibile dalle condizioni iniziali (praticamente impossibili da misurare in ogni dettaglio) ogni volta che vengono azionati.

5 Machine Learning

Il termine *machine learning* (apprendimento automatico) rappresenta una branca della ricerca sull'intelligenza artificiale volta ad individuare soluzioni algoritmiche in grado di migliorare attraverso l'esperienza. Tra questi gli algoritmi genetici e le reti neurali.

5.1 Gli algoritmi genetici

Gli algoritmi genetici, in breve AG, rappresentano un modello computazionale proposto nel 1975 da John Henry Holland (Fort Wayne, Indiana, 2 febbrario 1929 - Ann Arbor, Michigan 9 agosto 2015). Tale modello si ispira alle teorie evoluzionistiche traducendone le logiche in un modello informatico. In natura ogni individuo ha le proprie caratteristiche, specifiche proprietà visibili esternamente che ne costituiscono il fenotipo. Ed è tale fenotipo a definire le possibilità ed i limiti delle interazioni dell'individuo con l'ambiente circostante. Il fenotipo è a sua volta determinato dal genotipo, ovvero dall'invisibile patrimonio genetico del singolo. In prima approssimazione possiamo affermare che ad ogni gene corrisponde un caratteristico fenotipo. Grazie alla selezione naturale, premiando gli individui migliori, si favoriscono insieme anche i geni migliori. Oltre alla selezione naturale, l'altro principio cardine della teoria evoluzionistica è quello della variazione genetica, infatti se non vi fosse alcuna variazione genetica non avremmo alcuna possibilità di evoluzione. Le modificazioni genetiche avvengono per due vie: la prima è un processo combinatorio grazie all'apporto dei genitori; la seconda sono le mutazioni genetiche casuali. Queste producono nuovi geni che si possono trasmettere ai discendenti. Benché tali modifiche siano spesso molto limitate esse si accumulano nel corso di periodi molto lunghi determinando modifiche macroscopiche. I tempi di tali modificazioni non sono lineari e si possono documentare cambiamenti repentini in periodi di importanti rivoluzioni ambientali. Ad esempio in poche migliaia di anni l'uomo è riuscito a creare innumerevoli e molto diverse razze canine. Ancora, nel periodo della rivoluzione industriale una specifica specie di farfalle[72] divenne nera (prima era bianca) in quanto questo colore garantiva maggiori possibilità di mimesi negli ambienti impolverati dal carbone delle macchine a vapore.

Tali logiche evoluzionistiche sono quindi tradotte in linguaggi informatici ed utilizzate negli algoritmi genetici per la soluzione di molti problemi. In questo tipo di strumenti vengono create una popolazione di strutture che svolgono una certa azione in un ambiente (gli esempi di strutture vanno da un organismo vivente, ad un valore di una variabile indipendente di una funzione matematica). Gli algoritmi genetici richiedono, per poter funzionare, che vi sia un certo numero di queste strutture, ovvero una popolazione. Inoltre ogni struttura della popolazione deve essere codificata con una stringa di simboli di uno specifico alfabeto. Tornando agli esempi di prima, un organismo può essere codificato attraverso le basi nucleiche del DNA, un valore di una variabile da una stringa di valori binari.

Tutte le strutture vengono poi fatte funzionare nell'ambiente simulato e la performance di ogni singola stringa misurata con un valore che viene definito *fitness*. Inizialmente viene generata una certa popolazione producendo casualmente molte stringhe. Successivamente queste vengono fatte interagire con l'ambiente e viene valutata la *fitness* di ciascuna stringa. Alla prima popolazione se ne sostituisce quindi una seconda. Questa nuova popolazione verrà realizzata partendo dalle stringhe con migliore *fitness* passando attraverso una "mutazione" realizzata attraverso la combinazione di coppie di stringhe selezionate o attraverso il cambiamento casuale di alcune stringhe specifiche. Anche questa seconda generazione viene fatta interagire con l'ambiente misurandone poi le *fitness*. Tale processo viene iterato fino a che non se ne

[72] Biston betularia.

decide l'interruzione. Gli algoritmi generici utilizzano quindi le logiche evoluzionistiche per l'individuazione della "loro" miglior soluzione ad un problema dato.

E' interessante ed utile sottoporre agli algoritmi genetici il classico problema del commesso viaggiatore[73] che deve visitare tutte le città ma minimizzare i costi di spostamento e senza mai passare per la stessa città. Si da dunque la seguente matrice dei costi (i valori di *fitness* saranno inversamente proporzionali alla matrice dei costi):

	A	B	C	D	E	F
A	∞	5	7	4	8	6
B	5	∞	3	9	2	8
C	7	3	∞	4	7	6
D	4	9	4	∞	5	1
E	8	2	7	5	∞	3
F	6	8	6	1	3	∞

Viene costituita al tempo $t = 0$ la popolazione iniziale con i seguenti individui:

p_1=(A, B, C, D, E, F)

p_2=(B, C, A, F, D, E)

p_3=(C, B, F, A, D, E)

p_4=(F, E, A, B, D, C)

p_5=(B, F, E, A, C, D)

p_6=(A, C, F, B, D, E)

p_7=(A, B, C, D, F, E)

[73] Il problema del commesso viaggiatore è un problema matematico in cui l'agente venditore vuole visitare tutte le città in cui deve recarsi per lavoro facendo la minor strada possibile, senza mai passare due volte per la stessa città e tornando infine alla città di partenza.

con i relativi costi di percorso: c_1=26; c_2=24; c_3=29; c_4=31; c_5=35; c_6=43; c_7=20.

Si osserva quindi che gli individui con *fitness* migliore sono p_1 e p_7.

Una volta selezionati gli individui con *fitness* più alta possiamo procedere con delle mutazioni secondo regole prestabilite:

1. Togliere un elemento a caso e reinserirlo in un'altra posizione
   (C, A, F, B, E, D) → (C, F, B, E, A, D)
2. Togliere a caso una sottostringa e reinserirla in un'altra posizione
   (C, A, F, B, E, D) → (C, E, D, A, F, B)
3. Togliere a caso una sottostringa, invertirla e reinserirla nella stessa posizione
   (C, A, F, B, E, D) → (C, B, F, A, E, D)
4. ……

Omettendo alcuni passaggi possiamo osservare le varie popolazioni.

Viene quindi a formarsi la nuova generazione che verrà sottoposta al medesimo test per individuare nuovamente le stringhe migliori P(t) → P(t+1) ed anche in questo caso seguendo regole stabilite.

$t = 0$

p_1=(A, B, C, D, E, F)	c_1=26	p_5=(B, F, E, A, C, D)	c_5=35
p_2=(B, C, A, F, D, E)	c_2=24	p_6=(A, C, F, B, D, E)	c_6=43
p_3=(C, B, F, A, D, E)	c_3=29	p_7=(A, B, C, D, F, E)	c_7=20
p_4=(F, E, A, B, D, C)	c_4=31		

$t = 1$

p_1=(A, B, C, D, E, F)	c_1=26	p_5=(A, B, C, F, D, E)	c_5=24
p_2=(B, C, A, F, D, E)	c_2=24	p_6=(B, C, A, D, F, E)	c_6=20
p_3=(C, B, F, A, D, E)	c_3=29	p_7=(A, B, C, D, F, E)	c_7=20
p_4=(F, E, A, B, D, C)	c_4=31		

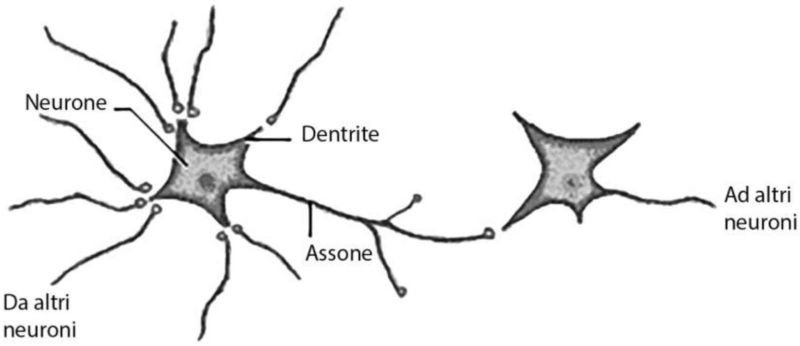
Il procedimento si interrompe all'avverarsi di una certa condizione che approssima la soluzione ottimale.

Gli algoritmi genetici sono sicuramente uno strumento interessante ed utile ma va precisato che la miglior soluzione individuata da questi procedimenti non sia necessariamente la soluzione ottimale.

Inoltre la numerosità della generazione, la frequenza di applicazione degli operatori genetici, e tutti gli altri paramenti sono variabili che possono essere determinati solo attraverso prove, valutando le performance e la qualità dei risultati ricevuti. Si rimanda quindi ad un supervisore umano che vigili sull'automatismo impostandone i parametri attraverso una valutazione dei risultati ottenuti rispetto a quelli attesi che devono logicamente precedere l'algoritmo.

5.2 Reti neurali

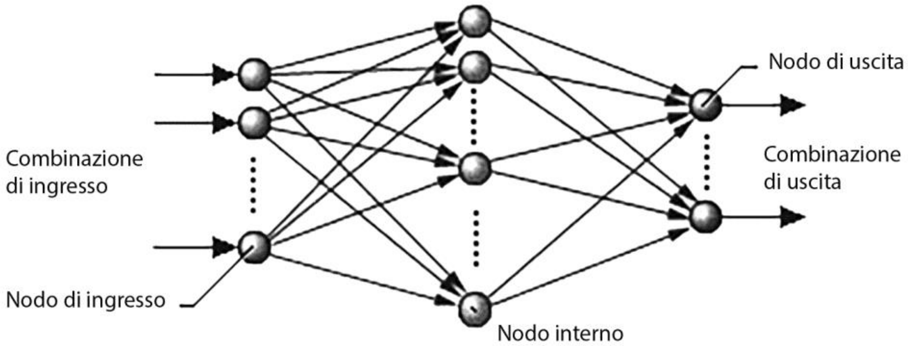
In natura, nel cervello umano, si stimano vi siano circa 10 miliardi di neuroni ognuno dei quali connesso con 10.000 altri. Ogni neurone è interconnesso con altri neuroni attraverso segnali chimici ed elettrici. I segnali in ingresso possono essere eccitatori o inibitori e se le somma di questi è tale da attivare il neurone questo trasmette il proprio segnale agli altri neuroni attraverso il proprio assone.



Lo studio delle reti neurali iniziò negli anni quaranta dal lavoro di Warren Sturgis McCulloch (Orange, New Jersey 16 novembre 1898 – Cambridge, Massachusetts 24 settembre 1969) e Walter Harry Pitts (Detroit, 23 aprile 1923 – Detroit, 14 maggio 1969). I due scienziati, il primo neurofisiologo, il secondo matematico, definirono il modello del neurone formale (1943).

In questo modello formale, gli input dei neuroni (in natura dendriti) sono definiti come O_j mentre l'insieme degli input che raggiungono il neurone sono $\sum_{j=1}^{n} O_j$. Ma gli input all'interno del neurone non assorbono tutti la stessa energia e quindi è necessario, anche nel modello formale, introdurre in fattore che "pesi" ogni singolo contributo. I vari pesi sono quindi definiti come W_{ij} e ne consegue che l'insieme degli input con relativi pesi nel neurone sarà $\sum_{j=1}^{n} O_j W_{ij}$. Più precisamente, dal momento che il neurone è definito come i l'espressione sopra indicata indica la sommatoria di tutti gli input del neurone ognuno dei quali viene "pesato" in relazione allo stesso neurone i. In uscita dal neurone formale, quelli che in biologia sono chiamati assoni, abbiamo gli output del neurone definiti come O_ie la relativa funzione di trasferimento $O_i = T(A_j)$ dove A_jè la funzione vista prima $\sum_{j=1}^{n} O_j W_{ij}$. La funzione di trasferimento ha l'importante compito di modellare l'output del neurone che sarà, a sua volta, l'input di altri neuroni. Aggiungiamo che il neurone accumula energia dai vari input e poi, raggiunta una certa soglia di attivazione, cede energia attraverso un suo unico output che, appunto, abbiamo detto essere la funzione di trasferimento.

Una volta definito il neurone formale possiamo passare alla rete neurale che è un insieme di neuroni connessi ognuno con i propri input e la propria funzione di trasferimento. Vi sono poi input e output esterni che interagiscono con la rete attraverso appositi nodi-neuroni detti di input o output.



In maniera più astratta la rete neurale può essere vista come una funzione che trasforma dei dati in ingresso fornendo dei dati in uscita: $f(x) = y$ dove x e y sono rispettivamente gli input e output della funzione (rete neurale).

Approfondendo le funzionalità delle connessioni tra i vari neuroni (le sinapsi) notiamo come queste variano nel tempo, o meglio, sia in natura che nella rete artificiale le energie di attivazione o i pesi non sono un dato costante ma mutano con l'apprendimento.

L'apprendimento è appunto un elemento fondamentale della rete neurale in quanto è solo attraverso questa fase che la rete si adatta. Conferendo il giusto peso alle varie sinapsi, una volta istruita, essa diviene capace di fornire risposte (output) adeguati agli stimoli. Possiamo definire quest'apprendimento come quel processo attraverso il quale i parametri liberi della rete vengono adattati attraverso un processo di stimolazione dell'ambiente nel quale la rete stessa opera.

Nell'apprendimento supervisionato, ad esempio, si fornisce alla rete un insieme di input ai quali corrispondono output noti (training set). Analizzandoli, la rete apprende il nesso che li unisce. In tal modo impara a generalizzare, ossia a calcolare nuove associazioni corrette input-output processando input esterni al training set.

Man mano che la macchina elabora output, si procede a correggerla per migliorarne le risposte variando i pesi. Ovviamente, aumentano i pesi che determinano gli output corretti e diminuiscono quelli che generano valori non validi.

Il meccanismo di apprendimento supervisionato impiega quindi l'Error Back-Propagation, ma è molto importante l'esperienza dell'operatore che istruisce la rete.

Altra forma di apprendimento è quello detto non supervisionato. Ad esempio mostrando alla rete molte foto di gatti, questa diviene capace di riconoscerli. «Ma com'è mai possibile che la rete apprenda cosa sia un gatto senza saper nulla del mondo, tantomeno dei gatti?. Le immagini di gatti, di per sé, contengono schemi – ossia tutto ciò che viene riconosciuto come i musi, i baffi, le zampe dei gatti e così via – in una apparentemente inesauribile varietà di posizioni, colori, angolature. Ma ciò che in effetti la rete neurale individua sono correlazioni incredibilmente sofisticate e complesse tra le varie immagini»[74]. E tali correlazioni sono relazioni e "somiglianze" ad input digitali (0 o 1) dei pixel che compongono le immagini.

Dopo un addestramento effettuato mostrando alla rete milioni di immagini, la rete diviene capace di riconoscere questi schemi in immagini mai viste prima. Il sistema funziona e funziona bene tanto da risultare addirittura migliore dell'uomo in alcuni ambiti del riconoscimento visivo. Le reti neurali possono essere immaginate come le corde della chitarra che risuonano a frequenze stabilite,

---

[74] J. Kaplan, *Intelligenza artificiale*, Luiss university press, Roma 2018, pag 60.

reagendo a modelli arbitrariamente complessi che rappresentano il loro addestramento in input. Queste sono specchi della loro esperienza e non imparano (nel senso comune del termine) le cose ma solo rispondono a quegli stimoli che sono allineati alle proprietà del loro mondo artificialmente costruito.

Le reti neurali hanno innumerevoli vantaggi nella risoluzione di problemi complessi o in contesti nei quali i dati sono imprecisi. Inoltre possono adattarsi all'ambiente. Esempi di applicazioni concrete di questi strumenti le possiamo trovare in finanza (con numerose applicazioni: previsioni sull'andamento dei mercati inclusi quelli valutari, analisi del rischio di credito, analisi del portafoglio, etc.), riconoscimento ed elaborazione delle immagini, analisi del parlato e riconoscimento vocale, simulazione di sistemi biologici, diagnosi mediche, inclusi i referti di TAC e risonanze magnetiche, controllo di qualità su scala industriale, data mining, ….

Tra i problemi che questo tipo di calcolo computazionale pone, c'è il cosiddetto effetto black box. Con questo termine si fa riferimento al fatto che non è possibile entrare all'interno della rete neurali (così come avviene per gli algoritmi genetici) per analizzarne gli stadi di elaborazione. Si è limitati nell'osservazione dei risultati ma non si possono approfondire le ragioni che a questi hanno portato, come in una scatola nera che non si può aprire.

6 Breve storia della sfida tra macchine e uomini

Quello che ad oggi più si avvicina al concetto di intelligenza artificiale sono Deep Blue, Watson e Deep Mind.
Questi computer, i primi due dell'IBM, il terzo di Google, sono stati in grado di battere i campioni umani rispettivamente nel gioco degli scacchi, in un quiz a domande e nel Go[75].

6.1 Deep Blue

Deep Blue è un computer progettato dall'informatico cinese Feng-Hsiung Hsu e dal canadese Murray Campbell, due scienziati finanziati dall'IBM per creare un computer che potesse sconfiggere il campione mondiale di scacchi. Garry Kasparov, nato nel 1963 a Baku, nell'Azerbaigian, era il campione mondiale indiscusso degli scacchi. La sfida si svolse a New York (una prima sfida si era tenuta l'anno prima a Filadelfia con la vittoria di Kasparov), tra il 3 e l'11 maggio 1997. Sedendosi alla scacchiera il russo aveva dichiarato «Difenderò la razza umana», ma perse la competizione svoltasi sulla lunghezza di 6 partite. Interessante il fatto che il maestro di scacchi dichiarò sempre di esser stato truffato e che a suo dire nella stanza di Deep Blue era presente anche un uomo in quanto, sempre secondo Kasparov, alcune mosse non potevano essere che umane. Pare quindi la macchina abbia superato uno speciale test di Turing degli scacchi.

[75] Il gioco del Go ha origine in Cina nel VI secolo A.C. Questo è giocato da due giocatori che collocano alternativamente le proprie pedine di colore bianco o nero su di una scacchiera con griglia 19 x 19. Nonostante regole relativamente semplici il gioco si apre ad una notevole complessità tanto che si stima vi siano $2{,}08 \times 10^{170}$ possibili posizioni.

6.2 Watson

"Jeopardy" è un noto quiz televisivo americano. Si gioca in che in tre manches in cui tre concorrenti rispondono a domande su vari argomenti. Per ogni domanda a cui rispondono correttamente ricevono una certa somma di denaro. Colui che alla fine della puntata possiede più denaro vince e torna in veste di campione nella puntata successiva, dove dovrà sfidare altri due concorrenti. Dal successo della versione americana sono nati in altri paesi format simili tra cui l'italiano "Rischia Tutto".

La sfida di IBM era quella di costruire una macchina in grado di competere a Jeopardy contro un concorrente umano. La difficoltà principale era capire le domande e formulare una risposta adatta. Per fare ciò i normali algoritmi di ricerca testuale (quelli ancora ampiamente usati in programmi di video scrittura o di lettura) non sono adatti poiché essi cercano delle parole chiave nel testo, senza comprenderlo, e quindi non sono in grado di rielaborare le informazioni ottenute per formulare una risposta adeguata.

Nel febbraio 2011 Watson partecipa a tre puntate del programma battendo, tra gli altri, Brad Rutter e Ken Jennings, rispettivamente il concorrente che ha vinto più denaro nello show e quello che ha vinto più puntate.

Watson non è collegato a Internet, il che vuol dire che i programmatori hanno dovuto fargli divorare un'enorme quantità di documenti prima di metterlo in competizione.

Durante la sfida Watson dovette rispondere a domande quali:

- Quali sono gli unici due presidenti consecutivi degli Stati Uniti con il medesimo nome?
- Di che esploratore si è celebrato nel maggio 1898 il quattrocentesimo anniversario del suo arrivo in India?

6.3 Deep Mind

Altro segno del progresso delle macchine è quello rappresentato dalla vittoria di Deep Mind, super-computer di Google, nel gioco del Go contro il campione umano Lee Sedol a Seul il 15 marzo 2016. Il gioco del Go, popolarissimo in Cina e giocato da oltre duemila anni, è ritenuto molto più complesso degli scacchi visto il maggior numero di mosse valide.

Le vittorie delle macchine, ognuna nei propri settori, se da un lato rappresentano passi fondamentali verso l'intelligenza artificiale, da un altro lato testimoniano ancora la lontananza di queste dall'obiettivo. Infatti, ogni macchina eccelle in quello per la quale è stata sviluppata, e questo è tutto. In altre parole Watson e Deep Mind sono eccezionali giocatori di Jeopardy e Go, ma nulla più e certamente non comprendono le regole del gioco. Altra cosa è infatti la capacità in generale di giocare un gioco. In questo caso alla macchina sono fornite prima le regole del nuovo gioco; questa deve gestire queste informazioni, anche se semplici ed espresse in modo formale, per giocare. Allo stato attuale siamo ben lontani da un simile risultato.

7 Conclusioni

"Can machines think?" È questa la semplice domanda di Turing nel suo saggio Mind del 1950. L'interrogarsi sulla possibilità dell'intelligenza artificiale equivale, cambiando il punto di vista, a chiedersi se le capacità della mente umana siano "replicabili", ovvero se il cervello sia una macchina, per quanto complessa.

La risposta a tale domanda non è altrettanto semplice e coinvolge materie tra loro molto diverse quali matematica e logica, informatica, ricerca tecnologia e filosofia.

Un primo scoglio è rappresentato dal che cosa sia l'*intelligenza*. Evitando di darne una definizione formale Turing sostiene che una macchina (ad esempio) è intelligente se riesce ad imitare il comportamento umano. Il Turing Test in maniera molto pragmatica ritiene di poter discriminare un essere che pensa da uno che è incapace attraverso un colloquio. L'apparenza di un comportamento umano sarebbe sufficiente, per il matematico inglese, ad inferirne la capacità sottostante di pensiero. Può sembrare un metodo semplicistico che evita lo scoglio di una definizione (e delle conseguenti caratteristiche) dell'intelligenza, ma d'altra parte è quello che quotidianamente facciamo quando, dando per scontato di essere intelligenti, decidiamo se lo sono anche animali, computer, persone, …

Nessun computer è riuscito, per ora, a superare il test di Turing. La cosa curiosa è che le macchine sono molto migliori degli uomini in attività complesse come il gioco degli scacchi o il calcolo, ma sono ancora lontanissime da «essere gentile, pieno di risorse, bello, amichevole, avere iniziativa, avere senso dello humour, riconoscere ciò che è giusto e sbagliato, commettere errori, innamorarsi, gustare le fragole con la panna, far innamorare qualcuno, apprendere dall'esperienza, usare le parole nel modo appropriato, essere l'oggetto del proprio pensiero, esibire una diversità di comportamenti pari a quella di un essere umano, fare qualcosa di veramente nuovo»[76].

Ad oggi quindi non possiamo dire che le macchine siano intelligenti. Ma è un problema tecnologico e di conoscenza delle regole della nostra intelligenza o c'è un qualche cosa nell'uomo che è impossibile replicare in un sistema artificiale? C'è un qualche cosa di qualitativamente diverso rispetto alla mente umana?

I fautori dell'AI forte sostengono che non vi sia nulla di così speciale nella mente umana: ciò che in noi viene svolto da circa 10 miliardi di neuroni interconnessi con stimoli elettrici e chimici non è altro che l'esecuzione di un algoritmo, per quanto molto complesso ed in larga parte ignoto. Alcuni suggeriscono che l'intelligenza e le capacità superiori dell'uomo sarebbero caratteristiche emergenti dall'enorme complessità ed articolazione della rete neuronale sottostante, teoricamente replicabile.

Altri, quali Searle, negando la possibilità dell'AI sostengono vi sia una differenza qualitativa tra cervello ed elaboratori elettronici. Egli, con il celebre esperimento della stanza cinese, sostiene che i computer sono incapaci ad attribuire senso ai simboli. Il computer darà effettivamente la risposta corretta alle domande che gli son poste in mandarino (grazie all'elenco delle operazioni combinatorie possibili dei simboli in ingresso) ma senza comprendere alcun significato e senza la possibilità di imparare davvero la lingua straniera.

Spesso i teoremi di incompletezza di Gödel sono stati chiamati in causa per dimostrare la superiorità umana. Secondo queste letture dei risultati del matematico austriaco, la dimostrazione dei limiti dei sistemi assiomatici formali

[76] A. Turing, *Computing machinery and intelligence*, pubblicato su Mind, Vol. LIX Nr. 235, 1950.

confermerebbe i limiti dei calcolatori (che in effetti utilizzano sistemi formali) rispetto all'uomo. Tale ragionamento non è corretto in quanto i risultati di Gödel escludono la possibilità di una dimostrazione di coerenza che possa essere proiettata sulle deduzioni formali dell'aritmetica. La verità aritmetica evidenziata nei teoremi è stata stabilita non come deduzione formale dagli assiomi dell'aritmetica, ma mediante un ragionamento metamatematico.
Restando all'interno del sistema formale, quindi è impossibile stabilire verità che invece possono essere colte riuscendo a guardare dall'esterno il sistema, da un punto di vista superiore.
La possibilità di "scalare" il punto di vista non può essere quindi esclusa per una macchina e d'altra parte fino a che punto può l'uomo allargare il proprio orizzonte?

Inoltre i teoremi di Gödel possono essere applicati ai sistemi formali. Se la mente non può essere ridotta alle logiche di tale categoria di sistemi, ci si chiede se al contrario i computer debbano necessariamente essere un sistema assiomatico formale con le relative le regole ed i limiti dimostrati dai teoremi di incompletezza. Ritengo che gli elaboratori attuali siano assimilabili ad un sistema formale. Questa opinione contingente non esclude però la possibilità che vengano realizzate macchine con complessità e caratteristiche superiori capaci di oltrepassare i limiti dei sistemi formali: le strutture e la limitatezza delle possibilità dei computer di oggi non possono, a mio avviso, escludere la possibilità teorica dell' AI.

Penrose, dal suo punto di vista matematico e fisico, sostiene la tesi che il cervello utilizzi anche logiche non logaritmiche. Ma queste logiche computazionali sono le uniche utilizzate dai computer attuali e da queste premesse se ne deduce la superiorità della mente. Tale posizione non nega la possibilità teorica che altre tecnologie (i computer quantistici) possano replicare le capacità umane, ma solo che le macchine con le quali abbiamo famigliarità, per quanto possano evolversi, resteranno sempre un passo indietro agli uomini, e al nostro servizio.

In queste riflessioni è difficile restare oggettivi ed imparziali. Non da alcun fastidio che un ghepardo sia molto più veloce di noi o che un uccello sappia volare; sull'intelligenza siamo molto più gelosi ed ipotizzare che qualcosa possa eguagliare o addirittura superare la nostra capacità "migliore" ci turba istintivamente, quasi minasse la dignità umana.
Ma se la ragione è la nostra caratteristica più nobile, allora dovremmo provare a conoscerla di più - necessariamente con la ragione stessa - e riflettere sui suoi limiti e sulla possibilità della sua replicabilità artificiale.

Nelle ricerche svolte nel presente lavoro non è stata trovata alcuna motivazione capace di fondare la presunta superiorità della mente.

Qualcuno potrebbe obiettare utilizzando l'esempio di Wittgenstein di quell'uomo che un giorno si svegliò con la testa di leone. Dopo lo stupore iniziale «la prima cosa che suggerirei, sarebbe di chiamare un dottore e fargli esaminare il caso in modo scientifico e, se non fosse per non fargli male, vorrei che fosse vivisezionato»[77]. È evidente che con un tale approccio il miracolo scompare e miracoloso resta solo un aggettivo per definire le cose che la scienza non ha ancora spiegato.
Traslando l'esempio nel caso dell'intelligenza artificiale, i detrattori dell'AI forte, potrebbero obiettare che l'uomo ha qualche cosa di "magico" che va oltre la logica e le possibilità di comprensione umana e che quindi gli strumenti scientifici

[77] L. Wittgenstein, *Lezioni e conversazioni*, a cura di Michele Ranchetti, Adelphi Edizioni, Milano 2012, pag. 16.

(la logica) non potranno mai identificare quel qualcosa che fa la differenza rispetto alle macchine.
Ma un tale approccio che delega al necessariamente ignoto i motivi della superiorità della "ragione" contraddice la ragione stessa quale unico strumento, se pur limitato, di indagine della realtà. Pare quindi presentarsi un cortocircuito inaccettabile: se la ragione e la conoscenza per la conoscenza sono ciò che ci rende uomini, non possiamo poi voler fondare la ragione stessa sulla "magia" e su un qualche cosa di imperscrutabile. E non importa se ciò ci conduce anche al non escludere la replicabilità dell'intelligenza.

Anche il libero arbitrio è coinvolto nelle discussioni relative all'AI. Spinoza sosteneva che la libertà fosse solo un'illusione e che tutto, sentimenti ed emozioni inclusi, fosse causato. In quest'ipotesi anche la libertà del singolo scompare e con essa anche quel qualcosa (in questo caso il libero arbitrio) che poteva essere indicato come elemento differenziale rispetto alla macchina. Ma la nostra mente ci fa apparire questa ipotesi inaccettabile ed incredibile; al tempo stesso ci è difficile negare l'esistenza di cose senza cause.

Infine, se l'uomo è il risultato dell'evoluzione naturale, è meraviglioso come la natura sia riuscita a costruire casualmente ed in relativamente poco tempo[78] uno strumento – la mente – così potente e complesso. L'uomo ha raggiunto conquiste

---

[78] L'*Homo habilis* comparve sulla terra circa 2,5 milioni di anni fa come discendente degli australopitechi (scimmie del sud). Dall'*Homo habilis* si sono succedute molte specie, tutte estinte tranne l'*Homo sapiens* (l'uomo moderno) apparso circa 300.000 anni fa. L'uomo è una animale recente se si considera che, ad esempio, è stata accertata l'esistenza di squali (tra le specie più antiche) circa 450 milioni di anni fa.

incredibili nell'ambito tecnico scientifico ma manca ancora di comprendere a fondo quell'elemento, il cervello, che le ha rese possibili.

Alla domanda *Can machines think?* sembra non si sia ancora riusciti a rispondere. E forse non è neanche possibile rispondere visto che, ragionando sulle caratteristiche dell'intelligenza e sulla sua replicabilità, sembra riproporsi la situazione delle proposizioni indecidibili all'intero di un ragionamento ricorsivo oltre il quale non è possibile salire.

8 Bibliografia

G. Boole, *An investigation of the Laws of Thought on which are founded the mathematical Theories of logic and probabilities*, Walton & Maberly, London 1854; rist. Dover New York, 1958 [trad. It. *Indagine sulle leggi del pensiero su cui sono fondate le teorie matematiche della logica e della probabilità*, Eiunaudi, Torino, 1976]

G. Chaitin, *Alla ricerca di omega*, trad. it. S. Frediani, Adelphi, Milano 2007

M. Davis, *Il calcolatore universale*, trad. it. G. Rigamonti, Adelphi Edizioni, Milano 2018

R. Hofstadter, *Gödel, Escher, Bach: un'eterna Ghirlanda Brillante*, Adelphi, Milano 1984

J. Kaplan , *Intelligenza artificiale*, Luiss university press, Roma 2018

G. Leibniz , *La Monadologia* (1720) trad. dal francese di Marianna Bacinetti (1856)

G. Leibniz, *Mathematische Schriften*, a cura di C.I. Gerhardt, 7 voll., Berlin (Voll. I-II) e Hll (voll. III-VII), 1849-1863; rist. Anast. Georg Olms, Hildesheim, 1962

M. Nuzzetti, *Metalogicon, International Journal of Pure and Applied Logic, Linguistics and Philosophy, 1993 VI, 2,* disponibile on line

J. McCarthy, M. L. Minsky, N. Rochester, C.E. Shannon, *A proposal for the Darmouth summer research project on artificial intelligence*, 1955

E. Nagel, J. Newman, *La prova di Gödel*, trad. it L. Bianchi, Bollati Boringhieri, Torino 2016

R. Penrse, *La mente nuova dell'imperatore*, trad. it. L. Sosio, Sansoni Editore, Roma 1997

H. Putnam, *Brains in a Vat*, 1981

S. Russell, P. Norvig, *Intelligenza artificiale*, Pearson Italia, Milano – Torino 2010

S. Russell, *Rationality and intelligence*, 1977, versione dell'autore disponibile online

M. Sacchetto *Putnam: logica, filosofia del linguaggio e riflessione etica,* in G. Fornero e S. Tassinari (a cura di), *le filosofie del Novecento***,** Milano, Bruno Mondadori 2002

B. Spinoza, *Etica*, trad. it. di Cristofolini, Edizioni ETS, Pisa 2014

A. Turing, *Computing machinery and intelligence*, pubblicato su Mind, Vol. LIX Nr. 235, 1950

A. Turing, *Intelligenza meccanica*, trad. it. G. Lolli, Bollti Boringhieri, Torino 2018

J. Searle, *La mente è un programma?,* tratto dalla rivista Le Scienze n. 259, marzo 1990

L. Vinciguerra, *Spinoza*, Carocci Editore, Roma 2016

L. Wittgenstein, *Lezioni e conversazioni*, a cura di Michele Ranchetti, Adelphi Edizioni, Milano 2012

E. Zalta, U. Nodelman, C. Allen, R. Anderson, *Alan Turing*, Stanford Encyclopedia of Philosophy, paper disponibile on line, 2013

E. Zalta, U. Nodelman, C. Allen, R. Anderson, *Artificial Intelligence*, Stanford Encyclopedia of Philosophy, paper disponibile on line, 2018

E. Zalta, U. Nodelman, C. Allen, R. Anderson, *Logic and Artificial Intelligence*, Stanford Encyclopedia of Philosophy, paper disponibile on line, 2018

E. Zalta, U. Nodelman, C. Allen, R. Anderson, *The Chinese Room Argument*, Stanford Encyclopedia of Philosophy, paper disponibile on line, 2014

E. Zalta, U. Nodelman, C. Allen, R. Anderson, *The Turing test*, Stanford Encyclopedia of Philosophy, paper disponibile on line, 2016