[INTRODUZIONE 3](#_Toc33739280)

[Capitolo 1. 4](#_Toc33739281)

[Un quadro generale sull’intelligenza artificiale 4](#_Toc33739282)

[1.1 Cos’è l’Intelligenza Artificiale 4](#_Toc33739283)

[1.2 Criteri per definire un’IA 7](#_Toc33739284)

[1.3 Il Ragionamento 9](#_Toc33739285)

[Capitolo 2. 11](#_Toc33739286)

[Dal pensiero alla macchina 11](#_Toc33739287)

[2.1 Prima di Aristotele 11](#_Toc33739288)

[2.1.1 i Presocratici 11](#_Toc33739289)

[2.1.2 Il concetto e la dialettica 12](#_Toc33739290)

[2.2 Aristotele e gli stoici: il Sillogismo 14](#_Toc33739291)

[2.3 Da Lullo a Pascal: i costruttori 18](#_Toc33739292)

[2.4 La Mathesis Universalis 21](#_Toc33739293)

[2.5 Leibniz: il sogno del calcolatore universale 25](#_Toc33739294)

[2.6 Dalla logica proposizionale alla logica dei predicati 28](#_Toc33739295)

[2.6.1. Boole 28](#_Toc33739296)

[2.6.2 Frege 31](#_Toc33739297)

[2.7 Russel e Whitehead: il logicismo e la Teoria dei tipi 35](#_Toc33739298)

[2.8 Hilbert e Brouwner: Formalismo vs. Intuizionismo 38](#_Toc33739299)

[2.9 Hilbert vs Gödel 40](#_Toc33739300)

[2.9.1 Entscheidungsproblem 40](#_Toc33739301)

[2.9.2 I teoremi di incompletezza 42](#_Toc33739302)

[2.10 Turing: la Macchina Universale 45](#_Toc33739303)

[Capitolo 3. 51](#_Toc33739304)

[Le macchine possono pensare? 51](#_Toc33739305)

[3.1 Il Test di Turing 51](#_Toc33739306)

[3.1.1 Le risposte di Turing a nove obiezioni 54](#_Toc33739307)

[3.2 Il funzionalismo computazionale 58](#_Toc33739308)

[**3**.2.1 Esperimento della protesi cerebrale 60](#_Toc33739309)

[3.3 Il naturalismo biologico di J. Searle 61](#_Toc33739310)

[3.4 IA forte e IA debole: la stanza cinese 65](#_Toc33739311)

# INTRODUZIONE

L’obiettivo di questa tesi di laurea è quello di ricercare l’origine filosofica delle concezioni fondamentali che hanno permesso lo sviluppo delle teorie che hanno portato alla costruzione dei calcolatori digitali e alla realizzazione di ciò che oggi viene conosciuto con il nome di Intelligenza Artificiale.

Nel primo capitolo, oltre ad alcuni cenni storici, ho cercato di delineare un quadro di definizione generale dell’IA, a partire dal significato di intelligenza. Dare una definizione di intelligenza è fondamentale per capire in quale direzione si è mossa la ricerca in IA al fine di trasferire questa facoltà alla macchina. Come vedremo, dall’amalgama di capacità attuabili da un’entità intelligente, ad emergere in questo caso sarà l’aspetto legato al raziocinio, e quindi al calcolo. Se per descrivere un essere intelligente non si può prescindere dalla sua capacità di ragionare, e se per ragionamento intendiamo un processo di pensiero rigidamente strutturato, allora per poter trasferire questo tipo di intelligenza alla macchina è necessario innanzitutto riuscire a formalizzare i meccanismi del ragionamento stesso.

Nel secondo capitolo ho rintracciato gli snodi principali della storia della formalizzazione del ragionamento logico.

# 

# Un quadro generale sull’intelligenza artificiale

## Cos’è l’Intelligenza Artificiale

Che cos’è l’intelligenza? Diverse teorie, nel corso dei secoli hanno cercato di dare una risposta esaustiva a questa domanda. In genere si ritiene che una condotta intelligente sia sinonimo di condotta razionale, tuttavia alcuni studiosi hanno allargato il campo rifiutando la definizione univoca di questa parola e ipotizzando l’esistenza di diversi tipi di intelligenze che, insieme, costituiscono la vasta gamma delle potenzialità umane. Potremmo dunque riferirci all’intelligenza come a quel complesso di facoltà mentali che consentono di pensare, comprendere o spiegare i fatti o le azioni, elaborare modelli astratti della realtà, intendere e farsi intendere dagli altri, giudicare, usare la logica, essere consapevoli, imparare, saper pianificare le proprie azioni, risolvere problemi, ma potremmo riassumere tutto ciò considerando in generale l’intelligenza come la capacità di acquisire ed elaborare le informazioni al fine di riuscire adeguatamente ad interagire ed adattarsi all’ambiente.

Il possesso di questa particolare facoltà ha fatto sì che la nostra specie, appunto l’homo sapiens, si differenziasse dalle altre specie in virtù della sua peculiare capacità di capire, comprendere, predire, manipolare il mondo circostante in modo cosciente, grazie alla capacità del pensiero.

Per migliaia di anni, suppongo in virtù della nostra coscienza, abbiamo cercato di comprendere questa nostra straordinaria capacità, ponendoci domande sull’origine e sul modo di funzionare delle nostre facoltà intellettuali: come pensiamo? Come funziona la nostra mente? ad oggi, le neuroscienze, che si occupano di descrivere il modo di funzionare del nostro cervello, non sono riuscite a fornirci una risposta precisa ed esauriente e l’origine del pensiero rimane per noi ancora un mistero.

Nonostante le enormi difficoltà tecniche e concettuali, sono state formulate alcune teorie della mente che, unite alla tecnica dell’ultimo secolo, hanno permesso di creare in principio macchine dotate di razionalità computazionale che, con il tempo e il progresso scientifico è andata ampliandosi fino a diventare ciò che oggi chiamiamo Intelligenza Artificiale (AI), cioè quel ramo della computer science che studia lo sviluppo di sistemi hardware e software dotati di specifiche capacità, tipiche dell’essere umano. In altri termini, l’IA è il campo di ricerca che studia la programmazione e progettazione di sistemi mirati a dotare le macchine di capacità prettamente umane. Le origini di questa disciplina risalgono al periodo immediatamente successivo alla Seconda Guerra Mondiale, la quale ha rappresentato un grande stimolo allo sviluppo tecnologico per via della necessità di sviluppare macchine in grado di raggiungere un’elevata potenza di calcolo da utilizzare nelle operazioni militari, fu in queste circostanze che nacque il Calcolatore Universale, o computatore. È interessante notare l’etimologia della parola che deriva dal latino *cum putare* e che riconduce a significati quali lo stimare, il reputare, il giudicare, il credere e anche il pensare.

La data di nascita ufficiale dell’IA risale all’estate del 1956, quando dieci massimi esperti di varie discipline (dalla matematica alla psicologia e dall’ingegneria alla neurologia) si riunirono al Dartmouth College, ad Hanover per tenere un convegno nel quale raccogliere molteplici contributi sul tema. L’obiettivo che gli studiosi si proponevano era quello di sviluppare uno studio che procedesse «sulla base dell’ipotesi che ogni aspetto dell’apprendimento e ogni caratteristica dell’intelligenza possano, in linea di principio, essere così precisamente descritti che una macchina possa simularli. Sarà fatto un tentativo di scoprire come costruire macchine che usino il linguaggio, formulino concetti e astrazioni, e risolvano problemi oggi di competenza solo dell’uomo, e siano capaci di auto-migliorarsi»[[1]](#footnote-1).

Durante questo convegno estivo, ci è noto che tra i partecipanti si accese un’intensa discussione proprio sul termine “Intelligenza Artificiale” proposto da John McCarthy, poiché ad alcuni di loro tale definizione sembrava essere inappropriata dato il significato equivoco della parola *intelligenza*. Simon e Newell, per esempio, continuarono a riferirsi al loro lavoro semplicemente come ad una complessa elaborazione dell’informazione.

Qual è dunque lo statuto epistemologico dell’IA? Come abbiamo visto, definirla risulta un compito arduo e potremmo darne varie definizioni, considerandola una tecnologia della conoscenza, come scienza generale del trattamento dell’informazione o ancora come una scienza alla cui base c’è una teoria dell’uomo e dei suoi processi cognitivi. Secondo alcuni, l’IA non può essere ridotta al semplice studio di computer sempre più sofisticati, ma riguarda l’uso di calcolatori e tecniche di programmazione al fine di far luce sui principi dell’intelligenza in generale e del pensiero umano in particolare[[2]](#footnote-2), altri sostengono invece che l’IA sia la scienza che si occupa di far fare semplicemente alle macchine cose che, se fatte dall’uomo richiederebbero intelligenza. Stando ad una definizione più tecnica, l’IA nella sua essenza consisterebbe nella ricerca del miglior programma agente per una specifica architettura, riducendo così il tutto ad un problema prettamente ingegneristico.[[3]](#footnote-3)

Qualsiasi sia la definizione che vogliamo dare all’IA, è lampante la sua trasversalità a molteplici discipline o meglio la sovrapposizione dei dominii di varie discipline quali ingegneria, fisica, informatica, linguistica, matematica, biologia, psicologia e non per ultima la filosofia, poiché la sfida di trasferire alle macchine l’intelligenza umana comporta necessariamente un fondamento teorico di uomo o di mente sul quale operare. Per questo motivo non esiste una definizione univoca di IA, almeno fin quando non si converrà in modo unanime sul significato di intelligenza, di mente, e quindi di uomo.

Il punto fondamentale della diatriba attorno a cui si accendono le polemiche, riguarda innanzitutto la possibilità della formalizzazione del comportamento intelligente, problema che affronterò in particolare nel terzo capitolo. È ovvio che dire se la creazione di un’IA sia realmente possibile dipende, come già detto, dalla definizione che si dà di intelligenza, ma dipende anche dalla possibilità di identificare quella componente esclusivamente associata all’essere umano come una caratteristica *formalizzabile e trasferibile* in un artefatto. I sostenitori dell’IA sono convinti che la formalizzazione del comportamento intelligente debba essere fattibile, tale convinzione sembra dipendere dal postulato ontologico che il mondo possa essere scomposto in elementi logici e dal postulato epistemologico che la nostra comprensione del mondo possa essere in seguito ricostruita rimettendo insieme questi elementi secondo delle regole. Questa convinzione, come vedremo nel secondo capitolo, è ben radicata nella storia della filosofia occidentale.

Per avere un quadro più chiaro di questa sfida così complessa, è necessario in questo caso entrare nell’ottica di una definizione degli scopi dell’IA.

## 1.2 Criteri per definire un’IA

In linea di principio ci sono due criteri per definire un IA[[4]](#footnote-4). Il primo misura il successo di un’IA in base alla somiglianza col comportamento umano, il secondo invece in base alla capacità di fare scelte razionali (quest’ultimo risponde al concetto ideale di intelligenza). In base allo scopo della creazione, i sistemi realizzabili in base ai criteri appena descritti sono i seguenti:

Sistemi che agiscono e pensano:

* Come esseri umani: a tal proposito è opportuno prendere in considerazione il Test di Turing, che ha fornito una definizione operativa dell’intelligenza. Turing suggerisce un test basato non sulle effettive doti intellettive della macchina ma semplicemente sull’impossibilità, per un umano, di distinguere il computer da entità umane. Per creare una macchina che ragioni come un essere umano, è necessario prima di tutto entrare dentro i meccanismi del nostro cervello per capire come pensiamo. A questo proposito è fondamentale sottolineare il contributo che la scienza cognitiva fornisce all’IA, poiché solo dopo che si è formulata una teoria della mente sufficientemente precisa, sarà possibile esprimerla sotto forma di un programma per computer, anche se come vedremo questo sarà possibile solo a patto di accettare teorie come quella funzionalista[[5]](#footnote-5) che ammettono che l’intero universo della mente umana venga ridotto ad elementi formalizzabili. Il lavoro in direzione della realizzazione di un’IA di questo tipo unisce modelli per computer e tecniche di sperimentazione psicologica nel tentativo di costruire teorie precise e verificabili sul funzionamento della mente umana;
* Razionalmente: affinché sia possibile creare un agente razionale, cioè un agente che fa la cosa giusta, è necessario essere in grado di rappresentare la conoscenza e applicarvi un ragionamento. Anche in questo caso l’enfasi è posta sulla correttezza delle inferenze. Tuttavia bisogna anche considerare che il corretto utilizzo dell’inferenza è solo uno dei molteplici meccanismi utilizzabili per arrivare alla razionalità[[6]](#footnote-6), infatti l’inferenza corretta non rappresenta *tutta* la razionalità, perché esistono delle situazioni in cui non è possibile determinare aprioristicamente un’azione giusta da fare. Inoltre è stato dimostrato che raggiungere la razionalità perfetta non è fattibile all’interno di sistemi elevatamente complicati in quanto i requisiti computazionali sono semplicemente troppo alti.

l tentativo di codificare formalmente il pensiero razionale risale agli antichi greci, ed in particolare ad Aristotele, che attraverso l’invenzione del sillogismo ha fornito un modello logico di deduzione che porta sempre a conclusioni valide. Si riteneva che queste leggi logiche fossero le stesse leggi del pensiero e che quindi governassero la nostra mente. Questa concezione è stata portata avanti nei secoli fino a quando i logici del 1800 sono riusciti finalmente a sviluppare una notazione precisa per formulare proposizioni riguardanti tutti gli aspetti del mondo (sempre nel limite della possibilità della loro formalizzazione) e le relazioni tra essi.

La tradizione logicista vuole raggiungere l’obiettivo di costruire sistemi intelligenti a partire dallo sviluppo di programmi siffatti, e nel prossimo capitolo vedremo nello specifico le tappe fondamentali dello sviluppo dei sistemi formali che, a partire dalle antiche intuizioni di Aristotele, si sono sviluppati fino a fornirci la base sulla quale è nata la moderna informatica e l’IA.

Aristotele definiva l’uomo come un animale dotato di ragione, ed a partire da allora la ragione è stata considerata l’essenza stessa dell’essere umano. Dal momento in cui abbiamo dichiarato la nascita dell’IA, abbiamo assistito al trionfo di una concezione speciale della ragione. Difatti se si può effettivamente dotare di ragione una macchina col fine di renderla simile ad un uomo, ecco che verrà confermata quella concezione dell’uomo che da duemila anni i pensatori dell’occidente cercano di sostenere, e che fondano proprio nella capacità di pensare in modo razionale. Ritengo essenziale a questo punto ricordare brevemente il pensiero di Heidegger, il quale considerava la cibernetica come l’apogeo della tradizione filosofica occidentale.

## 1.3 Il Ragionamento

Uno degli aspetti più interessanti dell’essere umano è la sua propria, peculiare capacità di ragionare. Il raziocinio che caratterizza la specie umana e la distingue dalle altre specie in quanto capacità essenziale, è una facoltà fondamentale che fin dai tempi antichi, nella cultura occidentale, è stata posta alla base della definizione stessa di *uomo*. Potremmo fare riferimento alle grandi religioni monoteiste e trovarvi sempre l’Uomo come essere distinto e posto al di sopra del resto del creato, uomo creato a immagine di Dio in virtù della sua capacità di intendere la suprema *ratio.* La filosofia stessa, nata intorno al VI-VI sec a. C. è sorta a causa della nostra intrinseca necessità di comprendere il reale, oltre l’analogia del mito, alla luce di un’indagine logico-razionale che potesse svelarne gli aspetti universali. La capacità di pensare, ragionare, astrarre è parte della natura più intima di ciò che Aristotele aveva definito come animale razionale, cioè come entità vivente dotata di un’anima superiore che rendeva possibile la capacità di una comprensione profonda e veritiera della realtà circostante e soprattutto della realtà non immediatamente visibile, quella raggiungibile attraverso la capacità di astrarre.

Ma cosa è, per definizione, un ragionamento? Un ragionamento è un processo di pensiero strutturato sulla base di regole definibili ed esplicabili in quanto universalmente valide (o almeno valide all’interno del sistema a cui si fa riferimento). La disciplina che si occupa dello studio delle leggi e delle funzioni che caratterizzano la struttura del pensiero in sé è la logica, che significa scienza del \greco{logos} λόγος, cioè scienza del pensiero. Individuare le tappe fondamentali dello sviluppo del pensiero logico è necessario per far luce su quei nodi cruciali che hanno condotto allo sviluppo di ciò che oggi chiamiamo Intelligenza Artificiale. Il tentativo di ridurre il ragionamento logico a una serie di regole formali risale ad Aristotele, ed era alla base del sogno leibniziano di un linguaggio computazionale universale, nonché del successo di Turing nel dimostrare che la sua macchina universale poteva svolgere qualsiasi calcolo. Tutt’oggi vi è un’interazione profonda che continua a crescere tra logica e informatica, la logica è infatti per l’informatica lo strumento fondamentale per costruire modelli. Calcolo e ragionamento logico sono due facce della stessa medaglia, e questa intuizione viene usata non soltanto per programmare i calcolatori in modo che sappiano eseguire un’impressionante varietà di compiti, ma anche per progettarli e costruirli. Potremmo considerare la stessa IA come il tentativo di concretizzare il postulato tradizionale che dichiara che la ragione umana sia scomponibile in operazioni governate da regole.

Uno dei primi logici della storia è senza dubbio Aristotele, al quale si attribuisce il merito di aver tentato di compiere per la prima volta una formalizzazione del ragionamento attraverso l’invenzione del sillogismo.

# 

# Dal pensiero alla macchina

## 2.1 Prima di Aristotele

### **2.1.1 i Presocratici**

Potremmo far risalire la storia dell’IA circa al VI secolo a.C., quando con Talete di Mileto ebbe inizio la trasposizione in ambito filosofico del modo di pensare matematico-razionale, pensiero che ebbe origine a partire dallo studio delle proporzioni fra le grandezze geometriche e astronomiche. Se nelle proporzioni o rapporti matematici si confrontano fra loro varie grandezze geometriche o fisiche, allo stesso modo nella filosofia, da Talete in poi, si iniziarono a confrontare fra loro, secondo criteri di necessità razionale, le ipotesi, le cause, le spiegazioni e le dimostrazioni relative a diversi fenomeni naturali e ai quesiti fondamentali dell'esistenza umana.

Ai fini dell’argomentazione di questa tesi non possiamo non ricordare la filosofia di Pitagora e della sua scuola, poiché Pitagora per primo ritenne che nei numeri vi fosse il principio e la spiegazione di tutti gli aspetti della realtà. La filosofia pitagorica identifica dunque nel numero il principio di tutte le cose. I pitagorici giungono a questa conclusione a partire dalla constatazione che tutti i fenomeni naturali (dalla musica alle stagioni, dai cicli astronomici ai cicli vitali) si realizzano con una certa regolarità, secondo rapporti calcolabili che fanno pensare a una loro dipendenza da principi numerici insiti in essi. I filosofi pitagorici però non hanno una chiara concezione della natura astratta dei numeri, ma ne hanno piuttosto un'idea fisico-geometrica, concepiscono, cioè, i numeri come un insieme di punti disposti nello spazio e raffigurati concretamente, infatti connettendo l'uno al punto, il due alla linea, il tre alla superficie e il quattro al solido, considerando cioè il numero come ente geometrico, è evidente come fosse possibile porlo alla base delle figure solide, cioè di ogni elemento della realtà. Su questa base i pitagorici possono definire il mondo come cosmo, cioè come un tutto ordinato, regolato da rapporti matematici. Se è vero che il numero è un archè di tipo fisico-naturale, di fatto esprime non semplicemente la sostanza di cui sono fatte le cose ma la loro struttura logica. Secondo questa visione pitagorica dunque, esistono delle leggi razionali che sono il fondamento sul quale si regge tutta la struttura della realtà, ciò significa che la realtà stessa potrebbe essere, in linea di principio, totalmente descrivibile, una volta comprese e descritte le leggi matematico-razionali del suo funzionamento.

### **2.1.2 Il concetto e la dialettica**

Se il problema principale da affrontare, quando si parla di IA, è quello attinente alla possibilità della formalizzazione della realtà e del pensiero, non possiamo non parlare di Socrate, che a detta di Aristotele è stato lo scopritore del *concetto,* ossia della definizione o conoscenza universale. Socrate, secondo la tesi platonica, avrebbe scoperto l’esigenza della definizione in risposta al relativismo linguistico adoperato dai sofisti, da qui la necessità di portare ordine nel discorso, prospettando la necessità di una precisazione, anche linguistica, dei concetti. Nei suoi dialoghi, Platone ci presenta un Socrate costantemente impegnato a far *generare delle definizioni* ai suoi interlocutori. La molla dell’intero processo è l’interrogativo definitorio simboleggiato dal *ti esti?* (Che cos’è?), ossia la continua richiesta di una definizione precisa di ciò di cui si sta parlando. Nell’Eutifrone, per esempio, Platone ci riporta l’immagine di Socrate intento a interrogare il suo compatriota che si appresta a denunciare di omicidio il suo stesso genitore in nome della pietà. «Io vorrei sapere» dice Socrate, «qual è l’elemento caratteristico di questa pietà che rende pie tutte le azioni, al quale io possa riferirmi, come regola in base alla quale giudicare le tue azioni e quelle degli altri uomini»[[7]](#footnote-7). Col senno di poi, possiamo individuare un’analogia tra la ricerca definitoria socratica e ciò che i moderni teorici dell’informatica chiamerebbero procedura operativa, cioè un insieme standardizzato di regole che possa dirci di momento in momento come agire.

Anche in Platone si rintraccia la necessità di stabilire definizioni stabili e universali, a differenza di Socrate però, egli le fonda nelle idee che sono modelli di realtà eterni e immutabili. Queste determinazioni universali e necessarie sono alla base della possibilità di elaborare delle definizioni esplicite che sono la condizione necessaria affinché possa essere possibile la conoscenza, al contrario, cioè in mancanza di queste definizioni esplicite, si tratterebbe non di un sapere ma solo di un opinare.

Nella Repubblica, Platone sostiene che la comprensione dipenda dalla Ragione, la quale implica un’analisi dialettica (anche se questa è rinforzata in ultimo dall’intuizione del significato dei concetti fondamentali utilizzati per la comprensione). L'origine di questo metodo nelle discussioni filosofiche è rintracciabile già in Zenone, il quale, sulla scia delle teorie parmenidee, sosteneva la tesi dell'immutabilità dell'Essere confutando le antitesi degli avversari tramite una dimostrazione per assurdo. Egli usava la dialettica come strumento di contrasto che approda indirettamente alla verità sulla base del principio di non contraddizione. Un metodo simile si ritrova ovviamente nei dialoghi platonici, dove Socrate cerca di mettere in evidenza le contraddizioni interne alle argomentazioni dell'interlocutore col fine di farlo cadere in contraddizione e in conclusioni assurde. Questo tipo di dialettica socratica è detto maieutica proprio perché l’interlocutore è costretto dalla forza della ragione e dai pungolamenti dell’argomentazione logica, a partorire da sé la verità. Come in Zenone, la dialettica non fa cogliere di per sé la verità, ma consente semmai di procedere alla confutazione degli errori e dei paradossi facendo uso della logica di non contraddizione.

Per Platone la dialettica è la scienza per eccellenza poiché permette la ricostruzione logica dei collegamenti che stanno a fondamento della realtà. Il procedimento dialettico definisce le relazioni esistenti nel molteplice rispecchiando la natura stessa delle idee. È da sottolineare, però, che in Platone le idee rimangono al di sopra della logica dialettica: esse sono accessibili soltanto per via di intuizione, e non sono dimostrabili, né ricavabili dall'esperienza sensibile.

La dialettica di Aristotele deriva da quella socratica e platonica, ma viene interpretata diversamente. Secondo Aristotele, le premesse su cui i suoi predecessori ragionavano erano principalmente le opinioni, emerse ed analizzate col metodo del dialogo, ed è a tal proposito egli distingue la dialettica dall'analitica, cioè dalla logica vera e propria, la quale studia la deduzione che parte da premesse vere per giungere a conclusioni logicamente fondate, cioè ad una dimostrazione. La dialettica, al contrario, ha per oggetto i ragionamenti che si riferiscono ad opinioni probabili, in questo senso la dialettica potrebbe essere definita come una logica dell'apparenza.

Nello stoicismo la dialettica verrà nuovamente identificata con la logica, come teoria dei segni che si riferisce alla realtà, respingendo la distinzione tra premesse vere e premesse probabili, diventando così la scienza del discutere rettamente, in conformità alle leggi universali del Lògos. Essa è scienza del vero e del falso, e quindi anche di ciò che non è né vero né falso, la logica verrà cioè viene intesa non solo in senso deduttivo, ma anche ipotetico, comportando un ampliamento di indagine del sillogismo aristotelico, come vedremo di seguito.

## 2.2 Aristotele e gli stoici: il Sillogismo

Aristotele assegna alla logica un ruolo fondamentale in filosofia in virtù della sua possibilità di garantire la correttezza del pensiero. Il termine logica tuttavia è di derivazione posteriore, il metodo di risoluzione del ragionamento attraverso le inferenze dei suoi elementi costitutivi era indicato da Aristotele con il termine *Analitica*, che deriva dal greco \greco{analisi} ἀνάλυσις e che non a caso vuol dire *scomporre, risolvere nei suoi elementi*.

Considerando uno dei princìpi fondamentali della logica aristotelica, cioè il principio del terzo escluso[[8]](#footnote-8), è possibile rintracciare un primo fondamentale collegamento con uno dei principi basilari del funzionamento dei calcolatori digitali. Il principio del terzo escluso mette in evidenza la natura binaria della logica aristotelica, che ammette solamente due possibili stati, l'uno il contrario dell'altro. A una logica del genere obbedisce anche il computer, il quale opera solamente su due stati: 0/1. In questa forma rigorosamente binaria i dati sono rappresentati ed elaborati.

Aristotele non si limita solo a riconoscere l’importanza della logica in quanto capace di garantire la validità dei ragionamenti, ma cerca per primo di *codificare* il pensiero razionale e formalizzarne il meccanismo formulando il *sillogismo aristotelico*: un metodo di deduzione di conclusioni a partire dalle premesse, compiendo il primo passo verso l’automazione del pensiero.

Secondo Aristotele, affermare o negare una proposizione non vuol dire ancora ragionare, il ragionamento sorge nel momento in cui si attua una connessione tra proposizioni stabilendo tra di loro un nesso inferenziale. Il ragionamento per eccellenza è formalizzato nel sillogismo, cioè un discorso nel quale sono poste alcune premesse e dalle quali si deduce necessariamente una conclusione per via del fatto che tali premesse sono state poste:

*A⇒B (Socrate è un uomo)*

*B⇒C (Tutti gli uomini sono mortali)*

*A⇒C (Socrate è mortale)*

Osservando la struttura del sillogismo riconosciamo un processo che individua concetti fondamentali mettendoli in relazione sulla base di leggi universali.

Secondo Aristotele, gli enunciati che compaiono nei sillogismi devono avere tutti la medesima struttura, devono cioè contenere un soggetto, rappresentato da un termine particolare, ad esempio *Socrate*, o universale, ad esempio *umano*, e un predicato.

La struttura proposizionale presa in considerazione da Aristotele, tuttavia, soffre di alcuni difetti. C’è innanzitutto da considerare la riduzione delle relazioni logiche alle relazioni grammaticali tra soggetto e predicato, prendiamo per esempio gli enunciati **1** e **2**:

1. *Luca* (sogg.) *mangia la mela* (predicato)
2. *La mela* (sogg.) *è mangiata da Luca* (predicato)

Secondo la logica aristotelica, se questi enunciati hanno una struttura sintattica diversa, ne risulteranno diverse anche le conseguenze logiche.

È evidente inoltre che in questo tipo di ragionamento, la validità della conclusione dipende dalla validità delle premesse iniziali, le quali, non essendo a loro volta dedotte, non sono necessariamente vere. Per poter porre rimedio a questa falla che avrebbe potuto inficiare la validità del ragionamento, gli stoici elaborarono un modello inferenziale chiamato *Modus tollendo tollens*. Questo modello sostituiva le proposizioni ipotetiche alle proposizioni assertorie di stampo aristotelico, ed aveva la seguente forma:

***Se*** *A* ***allora*** *B;*

***Se*** *non A,* ***allora*** *non B*

In questo modo la validità delle conclusioni non dipende più dalla verità delle premesse iniziali, le quali possono anche essere false, ma dipende esclusivamente dalla coerenza del ragionamento. In questa sorta di evoluzione del sillogismo aristotelico possiamo individuare il passaggio dalla *logica dei termini*, ad un primitivo modello di *logica proposizionale* in quanto l’inferenza è valida esclusivamente in virtù della forma dell’enunciato. La differenza sostanziale tra i due tipi di ragionamento sta nel fatto che la prima produce un ragionamento universalmente valido ma non necessariamente vero, in quanto la verità della conclusione dipende dalla verità delle premesse iniziali, che come abbiamo detto potrebbero anche essere false, tuttavia tale ragionamento soddisfa il requisito di coerenza. La logica proposizionale invece produce un ragionamento necessariamente vero, rispondendo dunque al requisito di verità, ma non universalmente valido in quanto è riferito ad un singolo individuo soltanto, come nel seguente esempio:

***Se*** *tutti gli uomini sono mortali*

***E*** *Socrate è un uomo*

***Allora*** *Socrate è mortale.*

Sarà poi Frege nel XIX sec., come vedremo, ad unificare logica dei termini e logica proposizionale elaborando un modello di ragionamento che potesse soddisfare allo stesso tempo i requisiti di verità e di coerenza.

L’invenzione del modello concettuale del sillogismo è in ogni caso di straordinaria importanza perché una tale possibilità di manipolazione dei concetti funge da base per lo sviluppo di ciò che in seguito diventerà, a distanza di secoli, la capacità di manipolare in modo automatico le informazioni alla base di algoritmi e programmi che oggi girano nei nostri computer.

L’invenzione del *congegno logico[[9]](#footnote-9)* del sillogismo riduce il ragionamento ad uno schema semplice che usa simboli generici e le loro inferenze per giungere ad una conclusione, e le basi dell’intelligenza artificiale, almeno in principio, erano proprio queste: l’utilizzo di sistemi simbolici astratti generali e un meccanismo che permetta di giungere ad una conclusione. Osservando a posteriori il quadro generale della situazione, si potrebbe guardare alla logica in AI come estensione continua della tradizione della logica filosofica[[10]](#footnote-10). Ma facciamo un passo indietro. La grande sfida di meccanizzare il pensiero, prima di giungere agli straordinari esiti contemporanei, ha interessato molte altre menti geniali del passato.

## 2.3 Da Lullo a Pascal: i costruttori

Dalla premessa secondo cui il ragionamento umano ha una primitiva natura numerica nella nostra mente, nasce il tentativo dell’uomo di costruire dei sistemi, o meglio delle macchine in grado di eseguire le operazioni logiche.

Ramon Llull è stato un intellettuale ispanico vissuto nel XIII secolo. A questo pensatore si deve uno dei primi tentativi di realizzare un metodo che potesse condurre, seguendo delle procedure formali, a delle conoscenze certe. Ciò egli che intendeva sviluppare era una metodologia intesa come

«scienza generale applicabile a tutte le conoscenze con dei principi generalissimi in cui è contenuto il principio delle scienze particolari come il particolare dell’universale»[[11]](#footnote-11).

Tale metodologia (*Ars magna*) finalizzata alla costruzione di una sorta di scienza del reale è esposta nell’opera che va sotto il nome di *Ars compendiosa inveniendi veritatem*, la quale può essere considerata come un vero e proprio trattato di logica. L’assunto di partenza del pensiero lulliano è che tutta la realtà, per quanto abbia una struttura complessa, possa essere sempre riducibile a una combinazione di termini semplici e determinabili. In seguito alla scomposizione della realtà in unità fondamentali, e successivamente alla riduzione di questi in lettere dell’alfabeto, Lullo dimostra che è possibile operare su tali termini per mezzo di un calcolo effettuabile su un congegno meccanico composto da ruote e figure mobili, che egli provvide a costruire concretamente, e che ruotando in base a delle precise regole razionali (*calculus ratiocinator*) avrebbe fornito una serie di molteplici combinazioni. La ruota di Lullo è una macchina logica che incarna materialmente, per la prima volta, il principio secondo il quale è possibile automatizzare il ragionamento. Affinché questo potesse avvenire occorreva che concetti complessi potessero scomporsi sino alla loro primitiva semplicità e che fosse possibile realizzare un linguaggio unico per tutte le scienze, una specie di *alfabeto del pensiero umano* composto da simboli (*characteristica universalis*) che potessero combinarsi tra loro così come avviene per le idee con l'attività di pensiero. Llull sviluppa dunque una metodologia utile, in linea di principio, a risolvere ogni problema attraverso la scomposizione di ogni quesito in parti più piccole e successivamente l'ulteriore riduzione di queste in lettere dell'alfabeto che fanno parte di ruote che saranno in grado di fornire infinite combinazioni, arrivando a descrivere non tanto una logica quanto una tecnica di ricerca. Lullo parte dal presupposto che ogni proposizione sia scomponibile in elementi costitutivi e che i termini complessi siano riducibili in molteplici termini semplici. Una volta descritti tutti i termini semplici della realtà, combinandoli in tutti i modi possibili si otterranno tutte le proposizioni vere possibili, con precisione matematica: nasce così il progetto teorico dell'arte combinatoria ripreso da Leibniz, e dal razionalismo cartesiano che ne individuerà però l'astrattezza. Dalla combinazione in tutti i modi possibili di questi simboli si otterranno tutte le proposizioni vere possibili: nasce così l'*Ars Magna,* che sarà poi ripresa in modo approfondito da Leibniz circa quattro secoli dopo, col nome di *Ars combinatoria* e che prevede una simbolizzazione del pensiero con cui operare calcoli logico-matematici.

L’automazione vera e propria della computazione proseguì nel 1500 con il progetto di Leonardo da Vinci di creare una macchina calcolatrice. La storia di questa presunta invenzione inizia con la riscoperta, avvenuta nel 1965, di due voluminosi codici leonardeschi nella Biblioteca nazionale di Spagna, oggi noti come Codex Madrid[[12]](#footnote-12), tra i fogli ritrovati ci sono giunti dei disegni progettuali di ciò che avrebbe dovuto essere, secondo le ricostruzioni, un calcolatore meccanico. Il progetto Leonardo da Vinci riporta la rappresentazione di un insieme di ingranaggi e numerose rotelle simili tra loro. La descrizione dell’uso cui sarebbe stato destinato lo strumento, fatta negli anni ’60 da un ingegnere che lavorava per l’IBM allo scopo di creare repliche dei macchinari leonardeschi, identificherebbe il macchinario come una primitiva calcolatrice meccanica. Nonostante da Vinci non ebbe modo di portare concretamente a termine il suo progetto, le moderne ricostruzioni ci dimostrano che esso era, in linea di principio corretto.

Anche se si ritiene che la prima calcolatrice mai costruita sia la famosa “pascalina” di Pascal, in realtà la prima vera e propria macchina calcolatrice fu realizzata attorno all’anno 1623 dallo scienziato tedesco Wilhelm Schickard. Solo qualche anno dopo, Pascal costruì la sua macchina ad ingranaggi che, a partire da una base in legno e una scocca in metallo contenente svariati meccanismi interni, era in grado di compiere le operazioni di addizione e sottrazione, basandosi come un abaco sul valore di posizione. È interessante ricordare, ai fini dell’argomentazione di questa tesi, di come Pascal si riferisse alla sua macchina in questi termini:

«Codesta macchina facilita ed elimina nelle sue operazioni tutto questo superfluo; il più ignorante vi trova altrettanto vantaggio che il più esperto; lo strumento supplisce al difetto dell’ignoranza o della scarsa abitudine e, attraverso movimenti necessari, fa da solo, senza neppure l’intenzione di colui che se ne serve, tutte le abbreviazioni possibili alla natura ogni volta che i numeri vi acconsentono. Tu sai parimenti come, operando mediante la penna, si sia in ogni momento obbligati a riportare o a prendere a prestito o a prendere in prestito i numeri necessari, e quanti errori si insinuino in questi riporti e prestiti a meno di una lunga abitudine, oltre ad una profonda attenzione che affatica ben presto la mente. Codesta macchina libera colui che se ne serve da tale vessazione; basta che egli abbia il giudizio ed essa lo libera dal difetto della memoria; e senza riporti né prestiti essa fa da sola ciò che egli desidera, senza che neanche ci pensi. […] Quanto alla comodità di questo movimento è sufficiente dire che è insensibile, andando da sinistra a destra ad imitazione della nostra maniera volgare [corsivi nostri] di scrivere […] La macchina calcolatrice produce effetti che si avvicinano al pensiero più di tutto ciò che fanno gli animali; ma essa non fa nulla che possa far dire che ha una volontà, come gli animali. »[[13]](#footnote-13)

Da questo frammento emerge la concezione pascaliana di una macchina calcolatrice vincolata al rigore del pensiero calcolante, essa si avvicina al pensiero degli uomini perché è in grado di compiere l’attività logica delle operazioni matematiche, ma Pascal in pochissime parole ci ricorda anche che non basta la sola capacità computazionale a descrivere un uomo.

## 2.4 La Mathesis Universalis

Aristotele, nella sua opera di logica formale (gli Analitici), aveva già intravisto la possibilità di una scienza unica, dove i concetti semplici potevano essere simboleggiati dalle lettere dell'alfabeto, e manipolati in modo da effettuare dei veri e propri calcoli.

L’arte lulliana sorta durante il Medioevo, fu dimenticata per un po’ di secoli, ma fu ripresa da Thomas Hobbes, che cercò poi di applicarla ad ogni campo del sapere. Hobbes afferma che la natura del ragionamento sia paragonabile ad un calcolo, la concezione che prevede l’equivalenza tra ragionamento e calcolo è quella su cui si fonda anche l’IA: «Quando uno ragiona non fa altro che ottenere una somma totale tramite una addizione di parti, o un resto sottraendo una somma da un'altra; il che se è fatto con le parole consiste nel ricavare dai nomi di tutte le parti il nome del tutto o dai nomi del tutto o da una singola parte il nome della parte rimanente. Sommando insieme due nomi si ha una affermazione, sommando due affermazioni si ha un sillogismo, sommando alcuni sillogismi si ha una dimostrazione; e dalla somma o conclusione di un sillogismo i logici sottraggono una proposizione per trovarne un'altra. Gli scrittori di politica sommando insieme i patti stipulati per trovare quali sono gli obblighi degli uomini, e i legislatori sommano le leggi e i patti per trovare che cos'è il diritto e che cos'è il torto nelle azioni dei privati. Insomma, in qualunque campo in cui c'è posto per l'addizione e la sottrazione c'è anche posto per la ragione; dove queste cose mancano la ragione non ha niente da fare»[[14]](#footnote-14).

L’idea che il ragionamento debba essere sottoposto alle leggi universali del calcolo è ripresa dal filosofo francese Renè Descartes, meglio noto come Cartesio. Cartesio è ritenuto il fondatore del Razionalismo: un movimento filosofico del ‘600 che esalta la ragione come unico strumento in grado di condurci ad una reale conoscenza della realtà.

Si definiscono razionalisti quei sistemi filosofici in cui la realtà è vista come governata da una serie di leggi e principi che sono perfettamente comprensibili con la ragione umana e che coincidono con il pensiero stesso[[15]](#footnote-15).

In generale i filosofi razionalisti sostengono che, partendo da principi fondamentali, individuabili intuitivamente o sperimentalmente, come gli assiomi della geometria, i principi della meccanica e della fisica, si possa arrivare tramite un processo deduttivo ad ogni altra forma di conoscenza. Come vedremo, questa concezione sarà sviluppata in seguito e svolgerà un ruolo cardine nelle teorie dei logici dell’ottocento e soprattutto nella teoria formalista.

Ai fini dell’argomentazione di questa tesi è necessario tenere in considerazione la visione razionalista perché fonda, in un certo senso, la possibilità di una comprensione profonda e veritiera della realtà da parte dell’uomo, proprio attraverso la facoltà della ragione, che diventa strumento di verità. Col razionalismo si afferma non solo che il mondo è comprensibile mediante delle leggi, ma soprattutto che l’uomo è in grado di entrare in contatto diretto con la realtà e conoscerla attraverso l’utilizzo rigoroso della sola ragione calcolante.

Cartesio individua gli elementi costitutivi della conoscenza umana nell’*intuitus* e nella *deductio*, ciò che tramite essi viene conosciuto necessiterà di un’elaborazione sistematica, le cui regole e procedure sono stabilite nella cosiddetta *Mathesis universalis*, o nel più famoso Metodo, il quale si occupa di stabilire una procedura universalmente valida del conoscere.

Nelle *Regulae ad diretione ingenii*, ricorrono due definizioni di metodo, l’una impostata sul concetto di regola o norma:

«per metodo intendo delle regole certe e facili, osservando le quali fedelmente non si supporrà mai come vero ciò che è falso, e senza inutili sforzi da parte della mente, ma con un graduale e continuo progresso della scienza si perverrà alla vera conoscenza di tutte le cose di cui si è capaci»[[16]](#footnote-16)

L’altra, che corrisponde alla quinta Regola, la quale afferma che:

«tutto il metodo consiste nell’ordine e nella disposizione di quelle cose alle quali deve essere rivolta l’acutezza della mente, per scoprire qualche verità. E lo avremo osservato con esattezza, se ridurremo gradualmente le proposizioni involute e oscure ad altre più semplici, e in seguito, con l’intuizione delle più semplici di tutte, tenteremo di salire per gradi alla conoscenza di tutte le altre»[[17]](#footnote-17).

In realtà queste due definizioni sono complementari perché l’ordine prescritto nella seconda esplicita la natura del graduale procedere secondo le regole della prima. La seconda in particolare è da intendersi come espressione della caratterizzazione del procedere naturale della mente nella sua naturale attività conoscitiva che procede secondo un processo che potremmo definire inferenziale.

La deduzione, secondo Cartesio, è l’atto con cui si conosce con certezza qualcosa che è di per sé non immediatamente evidente, a partire da altre conoscenze «con movimento continuo e mai interrotto del pensiero, che intuisce con evidenza ciascuna cosa; non diversamente da come conosciamo che l’ultimo anello della catena si connette col primo, anche se non abbracciamo in un’unica e medesima intuizione degli occhi tutti gli anelli intermedi da cui dipende quella connessione, pur che li abbiamo successivamente colti con lo sguardo uno dopo l’altro e ci ricordiamo che ognuno, dal primo all’ultimo, è connesso ai più vicini»[[18]](#footnote-18). Nell’oggetto conosciuto mediante la deduzione viene meno l’evidenza, ma è mantenuta la certezza nella successione dei passaggi proprio in quanto l’attenzione intellettuale viene tenuta fissa sul passaggio stesso, da un oggetto all’altro, anziché su un singolo oggetto come avviene nell’intuizione.

Con Cartesio abbiamo dunque una formalizzazione del processo conoscitivo, che nelle *Regulae* si esplicita nel concetto di *Mathesis universalis*. La *Mathesis* è infatti ciò che accomuna tutte le scienze, in virtù della sua funzione regolativa nei confronti di qualunque disciplina. Nel 1620 Cartesio scrive: «ho cominciato a capire il fondamento della scoperta mirabile. Si tratta dell’intuizione che tutte le scienze, al di là dei diversi oggetti specifici, sono unificate e unificabili in un fondamento comune, che è la Mathesis universalis, o disciplina universale dei rapporti formali secondo l’ordine e la misura»[[19]](#footnote-19).

Mathesis è dunque l’insieme dei principi non matematici grazie ai quali tutte le discipline, matematiche e non, sono sistematiche e complete, è la scienza generale che dà conto dell’ordine e della misura non riferita ad una materia specifica. I semi di tale Mathesis, dice Cartesio, sono rinvenibili negli antichi, per esempio nello stesso Aristotele, che già nel sillogismo aveva individuato *in nuce* la natura della Mathesis di cui parla Cartesio, caratterizzata come un insieme di princìpi capaci di organizzare qualunque sapere, e proprio in quanto tale è detta *universalis*.

Ma come si applica la Mathesis universalis ad ogni ambito del sapere? Bisogna anche qui ricondurci alla natura della ragione intesa come calcolo: l’aritmetica e la geometria si pongono come paradigma del processo conoscitivo, riportandoci alla necessità di seguire nella conoscenza, quel processo inferenziale e strutturato tipico delle scienze matematiche. Analogamente a quelle, la prescrizione fondamentale della VI regola impone infatti di disporre in serie le conoscenze in una disposizione gerarchica, non considerandole dal punto di vista del contenuto, ma formalmente, affinché «le une siano conosciute a partire dalle altre»[[20]](#footnote-20). Uno dei temi più noti delle Regulae è la *naturae simplices*, cioè la certezza che è possibile risolvere qualsiasi problema rintracciandone la struttura formale, grazie ad essa infatti è possibile escludere qualsiasi riferimento ontologico, istituendo una connessione delle conoscenze tra loro in funzione della derivabilità delle une dalle altre, in modo da raggiungere un primitivo conosciuto che Cartesio chiama semplice o assoluto. Le nature semplici sono colte mediante un’intuizione, e le proposizioni che le esprimono risultano essere quelle che fondano il sistema del conoscere. Potremmo paragonare queste intuizioni agli assiomi di Hilbert che vedremo nel paragrafo 2.8.

## 2.5 Leibniz: il sogno del calcolatore universale

Gottfried Leibniz è un filosofo e matematico tedesco vissuto nel XVII secolo, il suo talento emerse precocemente quando i suoi precettori gli presentarono il sistema logico di Aristotele elaborato nel IV sec a. C. che prevedeva la suddivisione dei concetti in categorie. L’approccio a questo antico tentativo di mettere in ordine la realtà contribuì a far nascere in lui l’idea che lo impegnerà per gran parte della sua vita, il suo obiettivo era sostanzialmente quello di ridurre il pensiero umano a unità fondamentali allo stesso modo in cui le parole possono essere scomposte nelle lettere che le compongono. Queste unità fondamentali del pensiero erano chiamate da Leibniz *characteristica universalis* e la manipolazione di tali unità o *alfabeto del pensiero* avrebbe permesso, secondo il filosofo, la possibilità di automatizzare il ragionamento: «Quando da ragazzo stavo imparando la logica, ed ero già solito indagare più profondamente le ragioni di ciò che mi veniva esposto, ponevo la questione ai miei precettori del perché, dal momento che si avevano le categorie (*praedicamenta*) dei termini incomplessi, mediante le quali vengono ordinate le nozioni, non vi fossero anche le categorie dei termini complessi, mediante le quali si potessero ordinare le verità; (…) Mi sembrava tuttavia che la questione in generale sarebbe stata in nostro potere se si fossero avute prima le categorie vere dei termini semplici e se, per ottenerle, si fosse costituito qualcosa di nuovo come un alfabeto del pensiero, ossia un catalogo dei generi sommi come *a, b, c, d, e, f*, dalla combinazione dei quali risultassero formate le combinazioni inferiori»[[21]](#footnote-21)

Il primo passo da compiere verso la creazione di un alfabeto dei concetti era proprio l’enumerazione di tutti i concetti e di tutte le combinazioni possibili di tali concetti. La difficoltà è evidente se si pensa alla mole immensa delle possibilità esistenti da tenere in considerazione se parliamo del pensiero umano, come è possibile individuare tutti gli elementi base dell’alfabeto del pensiero? Leibniz tratta il problema di come affrontare l’enorme numero delle disposizioni complesse di tutti gli elementi di base nella *Dissertatio de arte combinatoria* del 1666. La notazione da lui creata per il calcolo integrale e differenziale, utilizzata tutt’oggi, gli permetteva di eseguire facilmente calcoli dalla complessità molto elevata, come se fosse la notazione stessa a sbrigare tutto il lavoro, ma egli era animato da un’ambizione grandissima poiché non intendeva limitarsi al calcolo matematico in quanto era convinto che si potesse fare qualcosa di simile a anche per l’intera conoscenza umana, paragonando il ragionamento logico a un meccanismo, egli voleva trasferire la logica applicata al calcolo su proposizioni di altra natura, nelle quali non erano i numeri ad essere protagonisti ma il nostro pensiero, catalogato, ridotto ad unità di base in modo da poter essere manipolato alla stregua di qualsiasi elemento matematico, in virtù della sua possibilità di essere sottomesso alle medesime leggi logiche.

Leibniz sognava un linguaggio matematico artificiale universale in cui fosse possibile esprimere ogni minima sfaccettatura delle nostre conoscenze, un linguaggio che avrebbe permesso di definire, per mezzo di calcoli simbolici, le relazioni tra gli enunciati così da poter stabilire quali di questi erano veri o falsi, e regole di calcolo che mettessero in luce tutte le interrelazioni logiche esistenti tra le proposizioni del linguaggio stesso.

Leibniz non ebbe modo di assistere in vita alla realizzazione di questo progetto grandioso, riuscì però, nel 1673 a costruire un modello di macchina calcolatrice che, a differenza di quella costruita precedentemente da Pascal che era solo in grado di effettuare addizione e sottrazione, era capace di eseguire tutte le quattro operazioni fondamentali della matematica, poiché dotata di un meccanismo ingegnoso chiamato *ruota di Leibniz* spesso in uso ancora in pieno Novecento nelle macchine calcolatrici.

Nel 1674 egli descrisse addirittura una macchina in grado di risolvere equazioni algebriche, e la costruzione di queste macchine alimentò in lui la convinzione che potesse essere possibile la costruzione di un congegno capace di *ragionare* in senso generale. Bisogna però aspettare almeno fino al secolo successivo per avere i primi esemplari di macchine *programmabili*, in grado cioè di modificare il proprio lavoro in base alle istruzioni ricevute.[[22]](#footnote-22)

Le importantissime scoperte matematiche del Seicento, come la sistemazione della manipolazione delle espressioni algebriche, la scoperta cartesiana di poter ricondurre la geometria all’algebra, la sua stessa personale invenzione del calcolo infinitesimale e l’invenzione di un efficace simbolismo utilizzato nei calcoli di integrazione e derivazione, contribuirono ad alimentare in lui la convinzione che fosse fondamentale la scelta di simboli adatti e la ricerca di regole che ne governassero la manipolazione:

«ogni simbolo utilizzato non era un suono privo di significato ma stava per un concetto, e forniva un modello di quella meravigliosa idea che egli aveva in mente fin da ragazzo, di un alfabeto che rappresentasse tutti i concetti fondamentali».[[23]](#footnote-23)

Per mostrare quanto essenziale fosse per il pensiero deduttivo l’utilizzo di simboli ben scelti, Leibniz utilizza proprio l’algebra e non fa segreto della sua convinzione secondo la quale parte del segreto di questa disciplina risiedesse proprio nell’arte di usare correttamente le espressioni simboliche.[[24]](#footnote-24)

Tuttavia, per realizzare il suo programma di macchinazione del ragionamento umano, era necessario e fondamentale cercare dei simboli adeguati, da manipolare con quello che lui chiamava *calculus ratiocinator*, e che potessero rappresentare i mattoncini del pensiero, ma per fare ciò era necessario possedere la conoscenza umana in tutta la sua estensione. La sua pretenziosa ambizione di onniscienza è più o meno comprensibile se contestualizzata nella sua concezione del mondo che, essendo conforme a un piano perfettamente definito nella mente di Dio, non prevede nulla di indeterminato e quindi di indeterminabile. La volontà di assoggettare completamente il ragionamento umano alle leggi matematiche e deterministiche del calcolo, però, ci lascia intravedere una modello di uomo ambivalente, glorificato nella sua grandezza di animale razionale che è, allo stesso tempo, anche il suo limite.

A Leibniz si deve inoltre il merito di aver introdotto in occidente[[25]](#footnote-25) uno dei sistemi più importanti utilizzati per codificare informazioni: i numeri binari, in uso tutt’oggi nella scienza informatica. Questo sistema fu poi ripreso da George Boole.

## 2.6 Dalla logica proposizionale alla logica dei predicati

### **2.6.1. Boole**

Il contributo che George Boole, matematico e logico britannico vissuto nel XIX sec. fornisce alla storia della macchinazione del pensiero, consiste nell’aver matematizzato la logica, cioè l’aver fatto della matematica un metodo basato su simboli le cui leggi sono note e generali, creando una logica simbolica utilizzabile nella pratica, dando così vita a una parte del sogno leibniziano. Egli racconta di come, già all’epoca della scuola, gli fosse venuta in mente l’idea che dovesse essere possibile esprimere relazioni logiche in forma algebrica[[26]](#footnote-26), testimoniando una sorta di fede nell’idea leibniziana di poter creare un formalismo matematico capace di produrre da sé, automaticamente, la risposta corretta ai problemi. Boole procedette a un riesame della sillogistica, riprese l’antica logica aristotelica e stoica e ne fornì una nuova visione alla luce di ciò che era riuscito ad ottenere in campo matematico grazie all’introduzione di simboli e metodi algebrici.

La logica aristotelica operava su enunciati come *tutti gli uomini sono mortali*, a*lcuni uomini sono biondi*, *nessun gatto sa volare* e Boole capì che ai fini del ragionamento logico non era importante il significato di mortale o biondo, ma era importante il fatto che tale parola indicasse una classe di individui, fin qui nulla di nuovo. La novità fu introdotta nel momento in cui Boole indicò tali classi con delle lettere allo stesso modo in cui le utilizzava per rappresentare numeri e operatori in algebra. Arrivò a formulare un’algebra di tali classi, per esempio, se *x* e *y* stavano ad indicare due classi distinte, la notazione *xy* stava ad indicare l’insieme (oggi diremmo intersezione) degli oggetti presenti sia in *x* che in *y*, ma entriamo di più nel dettaglio:

«Assumendo la nozione di classe, siamo in grado di separare con un atto mentale, da una qualunque collezione concepibile di oggetti, quelli che appartengono alla data classe e di considerarli a parte dal resto. Possiamo concepire che tale atto di elezione, o un altro simile, sia ripetuto. Il gruppo di individui preso in considerazione può essere ulteriormente delimitato, selezionando mentalmente quelli tra di essi che appartengono, oltre che alla classe appena considerata, a qualche altra classe nota. E tale processo può esser ripetuto con altri elementi di distinzione, finché non arriviamo a un individuo che possiede tutti i caratteri distintivi che abbiamo preso in considerazione e che è un membro, al tempo stesso, di ogni classe che abbiamo enumerato. In effetti impieghiamo un metodo simile tutte le volte che, nel linguaggio comune, accumuliamo epiteti descrittivi per amore di una descrizione più precisa. Ora, le molteplici operazioni che abbiamo supposto di eseguire nel caso richiamato sopra sono soggette a leggi particolari (…) a qualcuno tali leggi appariranno così ovvie da essere annoverate tra le verità necessarie, e così poco importanti da non dover prestare loro speciale attenzione. (…) e se fossero diverse da come sono, l’intero meccanismo del ragionamento, anzi le autentiche leggi e la stessa struttura dell’intelletto umano, verrebbero modificate in modo essenziale. (..) si troverà che ogni proposizione logica, categorica o ipotetica, è suscettibile di espressione esatta e rigorosa. Ogni processo rappresenterà una deduzione, ogni conseguenza matematica esprimerà un’inferenza logica. La generalità del metodo ci permetterà di esprimere qualsiasi operazione dell’intelletto.»[[27]](#footnote-27)

La forza innovatrice del pensiero di Boole consiste nell’aver compreso l’autonomia del ragionamento simbolico-formale, infatti se la logica aristotelica si presentava sotto forma di una *lista di verità logiche*[[28]](#footnote-28), la logica booleana permetteva, grazie alla sua generalità, di creare *formule* applicabili in un numero elevatissimo di contesti, dalle normali interazioni umane ai ragionamenti più complessi, in quanto a differenza della logica aristotelica che operava su delle semplici proposizioni, la sua logica operava su proposizioni secondarie ben più complesse. Alla luce di ciò, il sillogismo aristotelico, che si occupa di una parte limitata di inferenze, viene ampiamente superato e inglobato in un metodo deduttivo molto più generale che trova la sua espressione nel calcolo algebrico.

Boole, inoltre, riprese da Leibniz l’interesse per il sistema binario che utilizzò per costruire la sua algebra, nella quale gli 0 e gli 1 assumevano rispettivamente valore di vero e falso in un sistema in cui, al posto delle operazioni aritmetiche erano utilizzati degli operatori logici come AND (congiunzione), OR (disgiunzione), NOT (negazione) i quali si riveleranno fondamentali nella programmazione e costruzione dei moderni computer. Senza saperlo infatti Boole aveva elaborato un’algebra capace di descrivere il funzionamento di circuiti elettrici dove 0 e 1 sarebbero stati utilizzati per descrivere i due stati fondamentali di acceso e spento.

Bisogna sottolineare come, per Boole, i principi e le leggi logiche che governano il nostro ragionamento, non siano generalizzazioni di osservazioni empiriche ma siano assolutamente necessarie, alla stregua delle leggi generali dell’aritmetica ed essendo tali, riflettono la struttura stessa della mente umana. Questo assunto è di fondamentale importanza perché ci permette di afferrare in pieno il filo conduttore che da Aristotele, passando per Leibniz, giunge fino al XIX secolo e che consiste nella comune pretesa di riuscire a formalizzare il pensiero umano.

### **2.6.2 Frege**

A Friedrich Ludwig Gottlob Frege, matematico vissuto tra il XIX e il XX sec. si deve il merito di aver realizzato il primo sistema logico pienamente sviluppato, capace di abbracciare tutti i ragionamenti deduttivi della matematica ordinaria.

Frege è considerato il padre del logicismo e il suo obiettivo primario era quello di fondare la matematica su basi logiche. Questa concezione della matematica come linguaggio formalizzabile sulla base di proposizioni analitiche si pone in netto contrasto con la visione kantiana che fondava la matematica sulle intuizioni pure interpretandola come una conoscenza sintetica a priori, per Frege al contrario, la matematica è una conoscenza analitica che si fonda sulla scienza logica. Per procedere a questa fondazione, però, Frege si scontra con un problema di non poco conto che consiste nel rischio che il ragionamento venga inficiato dalla limitatezza e dall’equivocità del linguaggio naturale, ed è proprio dalla necessità di superare questo ostacolo che nasce quell’esigenza di creare un linguaggio formale, generalissimo e privo di ambiguità attraverso lo sviluppo di una particolare simbologia che, in un certo senso richiama alla mente la *characteristica universalis* di impronta leibniziana[[29]](#footnote-29):

«Per evitare che in questo tentativo si introducesse inavvertitamente alcunché di intuitivo, tutto doveva svolgersi senza la minima lacuna entro la catena deduttiva. Cercando di soddisfare nel modo più rigoroso a questa esigenza, incontrai un ostacolo nell’inadeguatezza della lingua: infatti, malgrado la crescente pesantezza d’espressione, la lingua tanto meno mi permetteva di raggiungere quella precisione che il mio intento esigeva»[[30]](#footnote-30)

Il problema dello sviluppo di un linguaggio adeguato è affrontato nell’*Ideografia* del 1879, il cui titolo originale è *Begriffsschrift*, termine che la dice lunga sul contenuto dell’opera, costruito da Frege a partire dalla parola *Begriff* che significa «concetto» e *Schrift* che vuol dire «scrittura», con il sottotitolo *Linguaggio in formule del pensiero puro modellato su quello dell’aritmetica*.

Nell’*Ideografia,* che non fu accolta con grande entusiasmo dai suoi contemporanei, è possibile rintracciare la prima e rigorosa sistematizzazione formale della logica sia proposizionale che dei predicati, che scardina in modo definitivo il vecchio paradigma logico aristotelico. In sostanza Frege riformula il modello della logica tradizionale basato sullo schema soggetto-predicato, sostituendo a quello un modello aritmetico basato sullo schema funzione-argomento, gettando così le basi della moderna logica matematica.

Se Boole aveva compiuto dei passi avanti rispetto alla logica classica individuando correlazioni tra proposizioni che, a differenza di quanto faceva Aristotele, potevano essere espresse a loro volta da altre proposizioni secondarie, Frege comprese che queste correlazioni potevano essere usate anche per analizzare la struttura di una singola proposizione e ne fece il fondamento della sua logica.

La principale novità della logica fregeana, come appena accennato, consisteva nella considerazione della proposizione secondo la dicotomia di funzione-argomentoe nell’introduzione dei quantificatori.

Abbiamo già detto che la proposizione aristotelica era composta da soggetto e predicato, ma purtroppo questa dicotomia non riusciva ad esprimere enunciati più complessi e particolareggiati, come nel caso delle espressioni in cui sono presenti dei quantificatori quali *tutti, alcuni, nessuno* all’interno del predicato, la proposta fregeana invece riesce a risolvere anche il problema della logica aristotelica di come rendere una quantificazione interna a un predicato, facciamo qualche esempio per comprendere meglio considerando quest’enunciato:

1. *Alcuni studenti di filosofia* (A) *studiano qualche libro di matematica* (B).

Analizzando la proposizione da un punto di vista logico aristotelico, ne verrà fuori che tutte le A sono B, ma ciò risulta in contrasto col significato della frase stessa.

La dicotomia soggetto-predicato della proposizione aristotelica risulta inoltre limitante perché non permette di esprimere predicati relazionali. Se per esempio dicessi:

1. [*8* (A) ] *è* [*maggiore di 7*(B) ]

dall’analisi aristotelica ne risulterebbe semplicemente che A è B, e verrebbe completamente ignorata la relazione che intercorre tra *8* e *7*.

Per far fronte al problema degli enunciati relazionali, Frege introduce la dicotomia *funzione-argomento* che recupera proprio dalla matematica, in questo modo, prendendo sempre in considerazione l’enunciato **2**, avremmo che *8* e *7* sono argomenti, mentre l’espressione *è maggiore di* risulterebbe come funzione.

Una funzione è infatti un’espressione definita dalla forma f(x) che contiene uno o più costituenti indeterminati (x) tali ché, quando a questi componenti si assegnano dei valori, essa acquista un senso e un significato definiti, possiamo fare un esempio di funzione proposizionale dicendo:

*Socrate è “x”*

Finché x resta indeterminato essa non è né vera né falsa, ma quando ad x si assegna un significato definito, diventa una proposizione vera o falsa: vera, se questo significato rientra nella classe dei predicati attribuibili a Socrate (uomo, filosofo, ateniese ecc.), falsa in tutti gli altri casi.

La proposizione *Tutti gli uomini sono mortali* può interpretarsi come:

1. derivata dall’aver saturato con l’argomento *uomini* la funzione proposizionale: Tutti gli *x* sono mortali;
2. derivata dall’aver saturato con l’argomento *mortali* la funzione proposizionale: Tutti gli uomini sono *y*;
3. derivata dall’aver saturato con gli argomenti *uomini* e *mortali* la funzione predicativa: Tutti gli *x* sono *Y*.
4. derivata dall’aver saturato con l’argomento *sono* la funzione predicativa “Tutti gli x … y”, dove lo stesso verbo *sono* è una variabile sostituibile con un verbo diverso, come ad esempio *cercano*, *mangiano*, *inseguono* ecc.

Una volta stabilito il significato (valore di verità) delle singole proposizioni formate solo da soggetto e predicato e non più scomponibili, è poi sempre possibile, mediante l’impiego dei *connettivi logici****,*** verificare il valore di verità delle proposizioni composte e dell’intero discorso. I connettivi logici sono i seguenti:

* negazione  **¬** (*non*)
* congiunzione **∧** (*et*)
* disgiunzione  **∨** (*vel/aut*)
* implicazione **⇒** (*se.. allora..*)
* equivalenza  **⇔** (*se e solo se*)

Frege introduce inoltre l’utilizzo di quantificatori grazie ai quali diventa possibile, partendo da un sistema di assiomi, verificare se un qualunque asserto è o non è coerente con esso. I quantificatori si distinguono in:

* ***quantificatore esistenziale***: specifica se una proprietà riguarda l’insieme degli elementi componenti una classe, ovvero *ogni* elemento di quella classe **[∀*x* P(x)];**
* ***quantificatore universale***: specifica se una proprietà riguarda solo alcuni elementi **[∃*x* P(*x*)].**

Le novità introdotte da Frege resero possibile presentare le inferenze logiche come operazioni puramente meccaniche condotte per mezzo di cosiddette *regole d’inferenza* che riguardavano solo le configurazioni in cui erano disposti i simboli, pertanto potremmo definire un sistema formale come un sistema dotato delle seguenti caratteristiche:

1. un ***alfabeto***, ovvero un insieme finito di simboli;
2. una ***grammatica***, cioè un insieme di regole che specifica quali sequenze finite di simboli sono *formule ben formate* e quali no, la grammatica deve essere basata su un algoritmo;
3. Un insieme di ***assiomi***, ovvero un sottoinsieme dell'insieme delle formule ben formate;
4. alcune ***regole di inferenza*** che permettono di dedurre teoremi dagli assiomi, associando formule ben formate ad altre formule ben formate.

Fu così che nacque il primo esempio formale di linguaggio artificiale dotato di una sintassi precisa; da questo punto di vista la *Begriffsschrift* fu l’antenata di tutti i linguaggi di programmazione comunemente usati al giorno d’oggi[[31]](#footnote-31). Il tentativo di Frege di creare un linguaggio artificiale dotato di regole grammaticali e sintattiche di una precisione così approfondita, era sicuramente stato incentivato dall’antico sogno leibniziano di creare un linguaggio formale e universale che contenesse e derivasse in maniera automatica tutte le verità della filosofia e della scienza. È necessario riconoscere a Frege il merito di aver introdotto delle innovazioni estremamente importanti che hanno dato origine a quell’enorme mole di ricerche che, indirettamente hanno condotto Alan Turing all’idea di un calcolatore generale

## 2.7 Russel e Whitehead: il logicismo e la Teoria dei tipi

La posizione sviluppata da Frege secondo cui per lo sviluppo dell’aritmetica e della matematica stessa, siano necessari e sufficienti solamente i concetti della logica, è detta logicismo. La visione logicista considera la matematica riducibile alle regole e ai concetti generali della logica, come se fosse nient’altro che un’applicazione particolare della logica stessa, a partire dalla quale è possibile dedurre concetti, leggi e teoremi sulla base di assiomi[[32]](#footnote-32). L’indirizzo logicista, che ha avuto una grande importanza sia per lo sviluppo di sistemi automatici di calcolo sia per una definizione e limitazione delle potenzialità del calcolo stesso, concentra la sua attenzione sul ruolo della definizione e in primo luogo, proprio per evitare ogni infiltrazione di abitudini di pensiero consolidate, “intuizioni” mal definite o assunzioni non esplicitate, punta alla costruzione di un linguaggio “asettico”, assolutamente simbolico.[[33]](#footnote-33)

I *Principia Mathematica*, opera scritta a quattro mani da Bertrad Russel e Alfred Nortyh Whitehead nasce dalla volontà di rilanciare il programma logicista nato con Frege, in seguito al fallimento di quest’ultimo con la scoperta dell’antinomia da parte di Russel stesso.

L’opera di Russel e Whitehead si propone l’obiettivo di tentare una sistematizzazione delle basi della matematica partendo da un insieme definito di assiomi e di regole logiche, come possiamo leggere nella prefazione dell’opera:

«The present work has two main objects. One of these, the proof that all pure mathematics deals exclusively with concepts definable in terms of a very small number of fundamental logical concepts, and that all its propositions are deducible from a very small number of fundamental logical principles.

(Il presente lavoro ha due oggetti principali. Uno di questi, la prova che tutta la matematica pura tratta esclusivamente concetti definibili in termini di un numero molto piccolo di concetti logici fondamentali e che tutte le sue proposizioni sono deducibili da un numero molto piccolo di principi logici fondamentali)[[34]](#footnote-34)»

I *Principia* dunque, perfettamente in linea con il proposito logicista, affrontano il medesimo problema trattato nell'opera di Frege, che però come sappiamo si era arenata in alcune contraddizioni. Quest’opera tenta di superarle grazie all’elaborazione di una nuova teoria chiamata teoria dei tipi. La teoria dei tipi è una teoria logico-matematica che trova oggi applicazione nell’informatica, e si occupa di classificare generiche entità, raggruppandole in collezioni chiamate tipi. Questa teoria ha trovato un significativo campo di applicazione soprattutto nell'ambito della progettazione dei linguaggi di programmazione, in sostanza un sistema di tipi divide i valori manipolati dai programmi in insiemi chiamati appunto tipi attraverso l’esecuzione di un'operazione chiamata assegnazione del tipo o tipizzazione. Tale assegnazione serve affinché certi determinati comportamenti del programma siano, o non siano, possibili in base al tipo dei valori coinvolti in questi comportamenti.

Con un semplice esempio è possibile comprendere facilmente di cosa si sta parlando: supponiamo che un sistema di tipi classifichi il valore “casa” come stringa e il valore 8 come intero e sulla base di queste assegnazioni differenti, proibisca al programmatore di sommare l’elemento “casa” a 8. All'interno di questo sistema, non sarebbe assolutamente possibile, ad esempio, eseguire l'istruzione di programma “casa” + 8 in quanto elementi appartenenti a due differenti tipizzazioni.

Il vantaggio di questa "proibizione", ovvero dell'impossibilità di far eseguire al programma questa operazione, consiste nel fatto che non potrà mai capitare di sommare stringhe a numeri poiché sarebbe un’operazione che produrrebbe risultati privi di senso.

Il sistema di logica proposizionale formulato nei *Principia Mathematica*, può essere visto come un sistema di logica sentenziale costituito da un linguaggio e regole di inferenza[[35]](#footnote-35).

Nonostante l’opera di Whitehead e Russel aveva dato alla matematica un linguaggio artificiale che permetteva di affidare le dimostrazioni di teoremi a operazioni formali puramente simboliche. i *Principia* lasciavano però irrisolte alcune questioni, infatti non risolvono la questione riguardante le eventuali contraddizioni che potevano essere derivate dagli assiomi adottati, né tantomeno risolvevano il problema riguardante l’esistenza di verità matematiche che non possono essere provate o confutate nel sistema stesso. Le questioni sono state poi risolte dai teoremi di incompletezza formulati da Gödel nel 1931.

## 2.8 Hilbert e Brouwner: Formalismo vs. Intuizionismo

Un altro brillante logico matematico che contribuì alla successione di quegli studi che condussero alle intuizioni di Turing sul processo di calcolo, fu David Hilbert, anch’egli guidato da un profondo interesse per i fondamenti della matematica. Nel suo programma, matematica e logica sarebbero state sviluppate insieme come linguaggio simbolico puramente formale, che appariva come un sistema di formule e manipolazioni di simboli che potevano essere eseguite prescindendo da ogni significato, questo approccio è detto formalista. Diversamente dal logicismo che intendeva ridurre metafisicamente le entità della matematica alle entità della logica, il *formalismo* hilbertiano sostiene che gli enunciati matematici possono essere pensati come affermazioni intorno alle conseguenze di certe regole di manipolazione di elementi: in base alle regole logiche si possono derivare teoremi dagli assiomi e, di conseguenza, scoprire nuove proposizioni aritmetiche, ma il punto cruciale della visione formalista è che non importa affatto che le proposizioni derivate da una tale macchinazione siano vere nei termini di una corrispondenza metafisica con qualcosa di realmente esistente, potremmo pertanto dire che il sistema matematico deve la sua validità solamente dagli aspetti formali dei suoi teoremi e dei suoi processi intrinsechi, per cui un sistema matematico è valido se non presenta alcuna contraddizione al suo interno[[36]](#footnote-36). Per comprendere meglio questo concetto è necessario dare una definizione di sistema formale: potremmo definire un sistema formale come un sistema di assiomi dotato di regole di inferenza che consentono di volta in volta di generare nuovi teoremi.

Un sistema formale deve essere dotato di proposizioni di partenza in grado di definire astrattamente gli enti della teoria, gli assiomi A, e un insieme di regole di inferenza R. Sia A che R sono espressi in un linguaggio L sintatticamente preciso e a-semantico, un insieme di simboli e di operatori formali per la manipolazione di simboli. Un sistema formale deve possedere una serie di requisiti generali come coerenza o non-contraddittorietà, completezza sintattica (un sistema formale si dice completo quando, data una qualsiasi proposizione P, ben formata secondo le regole del linguaggio formale L, è possibile dimostrare che P può essere ricavata da A utilizzando le regole R), decidibilità (un sistema è decidibile se data una proposizione P è possibile dimostrare in un numero finito di passi se la proposizione appartiene o no al sistema utilizzando R)[[37]](#footnote-37). Inoltre un sistema formale può dirsi coerente se non esiste un'affermazione tale che l'affermazione stessa e la sua negazione siano entrambe derivabili nel sistema.

Hilbert cerca di dimostrare che i sistemi come quello dei *Principia Matematica* sono corretti e completi. Cercava in particolare una dimostrazione costruttiva, un *metodo effettivo*, e questo è quello che nel linguaggio moderno chiamiamo un algoritmo.

In questo modo, il lavoro fondamentale di un matematico, cioè dimostrare che una certa preposizione all'interno di una teoria è vera, quindi è un Teorema, o meno, sarebbe stato un lavoro del tutto meccanico.

In sostanza Hilbert intendeva utilizzare la matematica per giustificare la matematica stessa, si doveva cioè «dimostrare con la pura logica che i teoremi seguivano degli assiomi, senza l’influenza corruttrice di quello che possiamo “vedere” guardando una figura»[[38]](#footnote-38), assiomi per i quali non era possibile ricavare una dimostrazione. L’intento di Hilbert era dunque quello di fondare tutte le teorie matematiche su un insieme finito di assiomi che non conducevano a contraddizioni, comprimere cioè tutta la matematica all’interno di sistemi formali, fondando il cosiddetto metodo assiomatico. In tal modo, tutte le teorie complesse potevano essere fondate su teorie più semplici, retrocedendo in questo modo fino a fondare tutta la matematica sull’aritmetica, (ben presto però, Gödel dimostrerà, nel suo teorema di incompletezza, che non è possibile utilizzare l’aritmetica per dimostrare l’aritmetica stessa, perché i sistemi formali non sono completi).

Al polo opposto della teoria formalista, si pone l’intuizionismo. L’intuizionismo si fonda su alcuni punti essenziali circa lo statuto della matematica formulati dal matematico olandese Luitzen Egbertus Jan Brouwer. Innanzitutto egli considera ogni oggetto matematico non come qualcosa di esistente di per sé ma come un prodotto esclusivo dell’attività costruttiva della mente umana. Per l'intuizionismo la validità della matematica, in quanto scienza costruttiva, è indipendente dalla logica, proprio perché quest'ultima si riferisce a espressioni linguistiche, cioè a un momento che è successivo a quello del concepimento delle costruzioni mentali, quello della loro estrinsecazione verbale. Non è quindi possibile fondare la matematica sulla logica. La matematica quindi non è formale; i suoi oggetti sono semplicemente delle costruzioni mentali, e in quanto tali, la comunicazione di esse mediante il linguaggio può servire a suggerire ad altri costruzioni di pensiero analoghe alle proprie, ma non c'è alcuna garanzia che tali altre costruzioni siano le stesse. Ciò detto è facile intuire che secondo la visione intuizionista la matematica non dipende dalla logica; per contro, la logica è una componente della matematica.[[39]](#footnote-39)

## Hilbert vs Gödel

### **2.9.1 Entscheidungsproblem**

Quello di Hilbert può essere definito come un programma che aveva lo scopo di ridurre a processi algoritmici tutta la matematica. In un sistema del genere, in cui ogni proposizione deriva da una precedente ed ogni passaggio è perfettamente definito all’interno di un sistema formale, sarebbe stato possibile abbracciare l’intero ambito del sapere matematico senza eccezione alcuna.

Lo stesso Hilbert però, in occasione del Congresso dei matematici tenutosi a Parigi nel 1900 aveva posto un problema circa la possibilità dell’esistenza in matematica di alcune questioni non decidibili per via algoritmica, il cosiddetto *Entscheidungsproblem*, ovvero il problema della decisione.

Il problema consiste nel chiedere di eseguire meccanicamente (quindi tramite algoritmo) una procedura che per ogni formula espressa, dimostri che tale formula è o meno decidibile all’interno del sistema stesso. Il problema può considerarsi risolto se, per ogni proposizione dedotta nel sistema, si individuano una serie di operazioni e passaggi definiti e finiti che ne dimostrano la validità o la falsità[[40]](#footnote-40).

Hilbert era dell’idea che non ci fosse nulla che potesse sfuggire alla logica serrata dei suoi sistemi, come abbiamo visto, la sua costruzione della matematica all’interno di un sistema formale non ammetteva la possibilità dell’esistenza di proposizioni indecidibili, se tale sistema era definito come coerente e completo, sosteneva quindi che tutte le proposizioni potessero essere dimostrabili e affermava che la veridicità o falsità di tali proposizioni potesse essere deducibile nel sistema. I sistemi di procedure di calcolo destinati alla soluzione di problemi specifici, (cioè gli algoritmi), componevano buona parte del programma tradizionale degli studi matematici, Hilbert però cercava un algoritmo di un’ampiezza senza precedenti: l’algoritmo per l’*Entscheidungsproblem* avrebbe dovuto ridurre tutti i ragionamenti deduttivi umani a calcolo bruto, realizzando in buona misura il sogno di Leibniz. Secondo la visione hilbertiana, che parte dal presupposto che tutte le proposizioni matematiche sono derivabili da alcuni assiomi di partenza, se un sistema matematico contiene proposizioni indecidibili vuol dire che è basato su un insieme incompleto di assiomi e, quindi, nel caso della presenza di una proposizione apparentemente indecidibile, basterà semplicemente integrare quest’ultima tra gli altri assiomi![[41]](#footnote-41) Questa possibilità però venne presto negata ed Hilbert dovette ben presto ricredersi. Le pretese di onniscienza del sistema formalista vennero messe in crisi da un giovane matematico tedesco: Kurt Gödel.

### **2.9.2 I teoremi di incompletezza**

I due teoremi di incompletezza dimostrati da Kurt Gödel nel 1931 sono dei teoremi che mettono in luce i limiti dei sistemi formali. Secondo Gödel ogni sistema formale che abbia la complessità minima dell’aritmetica è necessariamente incompleto, ogni sistema assiomatico, cioè un insieme di proposizioni o principi che vengono assunti come veri perché ritenuti evidenti, che sia in grado di descrivere l'aritmetica dei numeri interi, ammette proposizioni logiche sugli interi che non possono essere dimostrate né confutate a partire dagli assiomi stessi, e a nulla potevano valere i tentativi Hilbertiani di integrare ulteriori assiomi. Il problema infatti, non deriva da un’incompletezza di assiomi, in quanto pur integrando assiomi su assiomi il problema dell’indecidibilità rimarrebbe invariato semplicemente perché ci sarebbero sempre, all’interno del sistema, alcune conseguenze logiche che non possono essere descritte da un algoritmo!

La risposta all’*Entscheidungsproblem* è quindi per Gödel negativa in quanto riesce a dimostrare l’incompletezza dei sistemi formali.

Il concetto fondamentale dei suoi teoremi di incompletezza è che per ogni sistema che formalizzi coerentemente l’aritmetica, è possibile costruire una proposizione sintatticamente corretta che non può essere né dimostrata né confutata all'interno dello stesso sistema, in altre parole, afferma che in ogni sistema formale coerente e assiomatizzabile si danno proposizioni indecidibili, cioè tali che né esse né le loro negazioni sono dimostrabili all'interno del sistema stesso. Questa affermazione va decisamente contro la pretesa logicista e formalista di poter ridurre la matematica ad una concatenazione di inferenze, dando così un contributo ampiamente significativo al dibattito che si teneva tra gli studiosi dell’epoca circa i fondamenti della matematica. L’importanza dell’intuizione gödeliana però va ben oltre, aprendo la strada a tutte quelle teorie matematiche e imprescindibilmente filosofiche che, attraverso tentativi di conferma o smentita del suo teorema, riguardano la possibilità di costruire un sistema (o programma, per utilizzare un termine proprio della scienza informatica) sulla base di algoritmi che possano descrivere in maniera assoluta il funzionamento di un’intelligenza in grado di riprodurre o simulare quella umana.

In particolare, se consideriamo il funzionamento della nostra mente analogamente a quello di un sistema formale, nei termini di inferenze logiche codificabili, è evidente che in seguito alla formulazione del primo teorema di Gödel, non sarà più possibile sostenere questo paragone.[[42]](#footnote-42)

Ma vediamo nello specifico di cosa si tratta. Il primo teorema di incompletezza di Gödel afferma che qualsiasi sistema formale coerente S entro il quale può essere eseguita una certa quantità di aritmetica elementare è incompleto. Questo significa che ci sono dichiarazioni del linguaggio di S che non possono né essere dimostrate né smentite in S stesso, questo equivale a dire che all’interno del sistema esistono delle proposizioni che sono indecidibili, rispondendo così all’Entscheidungsproblem proposto precedentemente da Hilbert.

Il teorema di Gödel non conferma semplicemente l'esistenza di tali affermazioni: il metodo della dimostrazione di Gödel produce esplicitamente una frase particolare che non è né dimostrabile né confutabile in S, e l'affermazione "indecidibile" può essere trovata meccanicamente da una specifica di S.[[43]](#footnote-43) La dimostrazione matematica effettuata da Gödel è molto tecnica e complessa, ma potremmo comprenderla attraverso un esempio relativamente semplice: Supponiamo che esista una proposizione P che dica "P non è dimostrabile in S", dove S è il sistema formale di riferimento. Assumiamo inoltre che ogni proposizione dimostrabile in S sia necessariamente vera. Ora, se P fosse dimostrabile in S, essa risulterebbe falsa perché andrebbe in contraddizione con il suo stesso contenuto, ma abbiamo detto che ogni proposizione dimostrabile in S non può essere falsa. Possiamo dunque dire che P è vera in quanto non può essere dimostrabile in S (quindi è indecidibile in S)[[44]](#footnote-44).

Questo teorema ci svela l’esistenza di una proposizione vera, ma la straordinarietà di questa dimostrazione sta nel fatto che la veridicità della proposizione P non è deducibile per via algoritmica. La sua indecidibilità all’interno del sistema fa sì che un computer, che lavora sulla base di algoritmi, non potrebbe mai dedurne la veridicità, al contrario di quanto è in grado di fare la mente umana, la quale è in grado di stabilire se quella proposizione è vera o no andando al di là di ogni possibilità di deduzione sistemica.

Ciò detto, potremmo affermare che l’intelletto umano è dotato di un surplus di capacità rispetto alla mera facoltà computazionale replicabile per mezzo degli algoritmi alla base dei programmi eseguibili dalle macchine, ma come potremmo definire questa sfuggente capacità? Intuizione?[[45]](#footnote-45) Oppure si tratta in ogni caso di un meccanismo il cui algoritmo ci è ancora sconosciuto?

In ogni caso, questo teorema ha ispirato il lavoro di Turing, che ne formulò una dimostrazione alternativa applicata all'informatica.

## 2.10 Turing: la Macchina Universale

La logica era stata, fino a questo momento, uno strumento fondamentale che forniva i componenti necessari per automatizzare una teoria del ragionamento attraverso il trattamento puramente formale, e quindi meccanico delle inferenze. L'idea di programmazione era dunque già stata inventata, a mancare era solamente una macchina che eseguisse tali operazioni. A progettare il primo calcolatore automatico di uso generale fu Charles Babbage nel 1834, ma a costruire il primo vero e proprio Calcolatore Universale fu Alan Turing, che ci arrivò in modo indiretto, a partire dai suoi studi matematici.

Dopo la formulazione del teorema di incompletezza di Gödel, era difficile pensare che potesse esistere un algoritmo come quello cercato da Hilbert. Ricordiamo che la questione era stata posta da Hilbert più o meno nei seguenti termini: esiste sempre, almeno in linea di principio, un metodo meccanico, algoritmico, rigoroso attraverso cui, dato un qualsiasi enunciato matematico, si possa stabilire se esso sia vero o falso? Un’ altra risposta negativa all’*Entscheidungsproblem* venne data da Alan Turing negli anni Trenta del ‘900.

Nella primavera del 1935, Turing si trovava a Cambridge ed ebbe la possibilità di frequentare un corso sui fondamenti della matematica, ed è proprio in seguito alla frequentazione di quel corso che egli decise di cimentarsi nell’impresa di dimostrare che un algoritmo come quello congetturato da Hilbert, non esisteva.

Turing, sfruttando il teorema dell'incompletezza di Gödel, dimostra l'impossibilità di ideare qualsiasi procedimento meccanico effettivo capace di stabilire in anticipo se una data proposizione si può dedurre logicamente da un assegnato sistema di assiomi. Il risultato di Turing stabilisce dunque dei limiti invalicabili al calcolo automatico.

Questo tentativo condusse indirettamente a fondare le basi della teoria della computazione, in quanto per dimostrare la sua tesi dovette inventare un modello matematico di calcolatore generale onnifunzionale (noto col nome di macchina di Turing) che fosse in grado di calcolare tutto ciò che è calcolabile mediante un processo algoritmico, tanto che oggi, per definire in modo formalmente preciso la nozione di algoritmo, si tende a ricondurlo alle elaborazioni effettuabili con macchine di Turing. Di conseguenza ne deriva che se si può dimostrare che un certo compito non può essere eseguito da una macchina di Turing, è certo che non esiste un processo algoritmico in grado di eseguirlo, e fu così che Turing dimostrò che non esisteva un algoritmo per l’*Entscheidungsproblem*, dimostrò cioè che nessuna macchina in grado di eseguire solo tali azioni di calcolo, poteva stabilire se una data conclusione era derivabile da premesse date usando le regole di Frege, infatti se di un qualche problema matematico si può dimostrare che è algoritmicamente insolubile, ne consegue che sarà insolubile lo stesso *Entscheidungsproblem*. Concluse così che l’algoritmo tanto agognato da Hilbert non esisteva.

Ma come si poteva realizzare una tale dimostrazione? Turing incominciò da una minuziosa e approfondita analisi del processo di calcolo, o meglio, incominciò proprio dal concetto di algoritmo, il quale consiste in un procedimento che risolve un determinato problema attraverso un numero finito di passi elementari, chiari e non ambigui.

Il metodo di istruzione del computer era molto importante per Turing. Far eseguire a un computer un compito particolare era possibile soltanto a patto che venisse effettuata un’adeguata suddivisione dell'istruzione in una serie di istruzioni più semplici, lo stesso processo che viene affrontato anche dai programmatori odierni. Turing era convinto che si potesse sviluppare un algoritmo per ogni problema. La parte più difficile stava nel determinare quali fossero i livelli semplici e come spezzettare i grossi problemi.

L’algoritmo è tipicamente definito da un elenco di regole che una persona può eseguire in modo meccanico e preciso. In particolare, un algoritmo deve avere queste cinque caratteristiche:

1. I passi che costituiscono lo schema devono essere *elementari*, ovvero non ulteriormente scomponibili;
2. I passi che costituiscono lo schema devono essere interpretabili in modo univoco dall’esecutore, sia esso umano o artificiale;
3. L’algoritmo deve essere finito, ossia composto da un numero definito di passi legati ad una quantità definita di dati in ingresso;
4. L’esecuzione dello schema deve avvenire entro un tempo finito;
5. L’esecuzione dello schema algoritmico deve condurre ad un unico risultato.

Eliminando ogni dettaglio inessenziale, Turing riuscì ad individuare le azioni di base necessarie per lo svolgimento di un calcolo. È importante premettere che una persona che esegua un calcolo è soggetta ad alcuni vincoli: innanzitutto in ogni stadio del calcolo l’attenzione deve essere rivolta solo a pochi simboli, e inoltre in ogni stadio l’azione intrapresa dipende solo da quei simboli su cui si focalizza la sua attenzione e dal suo stato mentale del momento.

Detto ciò, il calcolo può essere inteso come un processo con le seguenti caratteristiche:

1. viene eseguito scrivendo dei simboli nelle caselle di un nastro di carta;

2. a ogni passo la persona che esegue il calcolo fa attenzione al simbolo scritto in una sola di queste caselle;

3. l’azione successiva dipenderà da questo simbolo e dallo stato mentale della persona;

4. tale azione consisterà nello scrivere un simbolo nella casella osservata ed eventualmente nello spostare l’attenzione sulla casella immediatamente a destra o a sinistra.

Le analisi del processo di calcolo (cioè delle azioni fondamentali eseguite da un essere umano nel calcolo) portò alla conclusione che le medesime azioni potevano benissimo essere eseguite da una macchina capace di svolgere quelle stesse azioni di base, e fu così che Turing arrivò all’idea della sua macchina calcolatrice, cioè uno di quegli strumenti dalle caratteristiche rigidamente definite che oggi chiamiamo macchine di Turing[[46]](#footnote-46).

Poco più di un secolo prima, come precedentemente accennato, Charles Babbage aveva avuto l’idea di costruire una macchina calcolatrice automatica, la cosiddetta *macchina analitica* che avrebbe dovuto essere in grado di svolgere ogni tipo di calcolo numerico. Babbage tuttavia intendeva costruire una macchina usando solo componenti meccaniche, utilizzando schede perforate simili a quelle utilizzate nel telaio di Jacquard (tecnologia che verrà in ogni caso utilizzata nella costruzione dei primi calcolatori), tale macchina però era priva del meraviglioso apparato concettuale che caratterizzava la macchina di Turing.

A livello tecnico, la macchina di Turing si compone di un nastro sul quale i simboli scritti sono rappresentati da informazioni in codice. Il nastro si muove avanti e indietro, mentre agli stati mentali dell’operatore corrispondono differenti configurazioni delle componenti interne della macchina. Quest’ultima va progettata in modo da scandire, istante per istante, uno solo dei simboli del nastro e a seconda della sua configurazione interna e del simbolo scandito, scriverà sul nastro un certo simbolo che rimpiazzerà quello scandito precedentemente, dopodiché o continuerà a scandire la stessa casella o si sposterà di un passo verso destra o verso sinistra. Ai fini del calcolo però conta solo che la macchina abbia la capacità di assumere un certo numero di configurazioni, anche dette stati.

Che cosa è dunque necessario per descrivere una di queste macchine? Ci serve innanzitutto un elenco di tutti i suoi possibili stati, per ciascuno di essi e per ogni simbolo che possiamo trovare sul nastro è necessario specificare l’azione svolta dalla macchina quando è in quello stato e si trova davanti a quel simbolo, questa azione consiste semplicemente nell’eventuale sostituzione del simbolo nella casella scandita, nello spostamento di una casella verso destra o verso sinistra e in un eventuale cambio di stato secondo il seguente enunciato:

*quando la macchina si trova nello stato* ***R*** *e legge sul nastro il simbolo* ***a****, sostituisce* ***a*** *con* ***b*** *e passa nello stato* ***S***.

Questo enunciato può essere formulato in un’espressione chiamata quintupla, perché occorrono esattamente cinque simboli per descriverlo:

**R a : b → S**

Calcolare attraverso una macchina di Turing significa dunque spostarsi su questo nastro, avendo la possibilità di leggere quello che sta scritto nelle caselle e di stampare o cancellare simboli, passando successivamente in stati diversi. Quando la macchina di Turing esegue un calcolo, stando alla definizione precedentemente data di algoritmo, dopo un certo numero finito di passi si dovrebbe fermare. A questo proposito nasce un problema teorico molto importante, detto *problema della fermata*: la macchina di Turing si fermerà davvero tutte le volte che la sottoponiamo ad un processo di calcolo? La risposta è no. Come aveva già dimostrato Gödel, vi sono delle funzioni matematiche che non sono computabili in un numero finito di passi, secondo il funzionamento della macchina di Turing. Questo significa che in certi casi, la macchina di Turing potrebbe andare avanti all’infinito e non fermarsi mai. Turing aveva creato il modello matematico di un calcolatore generale per dimostrare che non si poteva usare una tale macchina per risolvere l’*Entscheidungsproblem*, ma per passare da questo risultato alle conclusioni che non esisteva nessun algoritmo, di nessun genere, capace di risolvere un tale problema, avrebbe dovuto dimostrare che la sua macchina era in grado di svolgere davvero *qualsiasi* calcolo.

Fu così che Alan Turing formulò il modello matematico di un calcolatore universale in grado di svolgere in modo automatico qualsiasi tipo di computazione, formulò cioè il modello di una macchina che fosse in grado, da sola, di svolgere tutti i compiti di qualunque macchina di Turing: la macchina di Turing universale.

In poche pagine di quella che oggi chiameremmo programmazione riuscì a descrivere come fosse possibile produrre le quintuple di una simile macchina universale che, partendo dal codice numerico di una macchina T, seguito sul nastro dal numero dato in ingresso a T, facesse esattamente ciò che avrebbe fatto T se le fosse stato fornito quell’ingresso, in pratica aveva creato il modello di una macchina che eseguisse da sola i compiti di molteplici macchine di Turing, più o meno come un programma è in grado di trattare altri programmi come dati, ed è proprio per questo motivo che tale macchina è detta universale.

La macchina universale di Turing può essere considerata come il primo esempio di programma interprete, in quanto il suo funzionamento si basa sull’interpretazione di una successione di quintuple analogamente a quanto succede nei moderni programmi scritti in un determinato linguaggio di programmazione che altro non è se non un insieme di dati che deve a sua volta essere interpretato.

Era da tempi remotissimi che si pensava alla realizzazione di macchine calcolatrici, ma è con Turing e con la sua macchina universale che per la prima volta l’uomo si affaccia finalmente alla possibilità concreta di riuscire a costruire un meccanismo in grado di poter calcolare tutto il calcolabile, riuscendo così finalmente a realizzare l’antico sogno di Leibniz e di chi prima di lui aveva mosso, più o meno inconsapevolmente, i primi passi verso quella che potremmo definire come *macchinazione del pensiero*. Sembrava che l’antica ambizione dei filosofi avesse trovato la tecnologia che le mancava: l’elaboratore, ultrarapido, universale e fornito di quelle regole ricercate lungo il corso dei secoli, che potevano permettergli di tradurre il ragionamento in calcolo.

# 

# Le macchine possono pensare?

## 3.1 Il Test di Turing

La definizione tecnica di IA, secondo la quale essa consisterebbe nella ricerca di un buon agente, data una certa architettura, si riduce ad un problema ingegneristico. Nell’ambito filosofico però, il problema è molto più ampio perché ci si trova davanti alla questione del confronto tra due architetture: quella delle macchine e quella umana.

Nell’articolo *Computing Machinery and Intelligence* pubblicato sulla rivista Mind nel 1950, Alan Turing affronta il problema di come riconoscere se una macchina sia dotata di capacità intellettuali equivalenti a quelle umane, e a tal proposito elabora un test, conosciuto col nome di Test di Turing. L’articolo del 1950 si apre con una domanda breve, concisa e fondamentale: «*can machines think?*»[[47]](#footnote-47) cioè, le macchine possono pensare? Il problema posto da Turing ha assunto una rilevanza profondissima per via della sua portata rivoluzionaria sia nell’ambito strettamente tecnico-scientifico, nell’ambito delle neuroscienze e soprattutto nell’ambito della filosofia[[48]](#footnote-48), la quale a partire da questo momento non si sarebbe più interrogata soltanto sulle capacità e sui limiti del pensiero umano, ma anche sul complesso rapporto tra macchina e pensiero.

Ovviamente il tema è molto controverso in quanto non è possibile definire in modo univoco che cosa sia l’intelligenza e il pensiero umano, secoli di teorie filosofiche ci hanno fornito solo arbitrarie interpretazioni e tutt’oggi gli studi neuroscientifici, oltre che a descriverci il funzionamento meramente chimico-fisico dei processi cerebrali, non riescono a fornirci una definizione certa e definitiva di cosa sia di preciso quella straordinaria capacità umana che ci distingue dagli esseri inanimati.

Turing tuttavia, per evitare di entrare nel merito di complesse questioni etiche e filosofiche derivanti dal significato arbitrario attribuito ai termini macchina e pensiero, decise di proporre la questione in un altro modo, suggerendo un test basato semplicemente sull’impossibilità di distinguere un computer da esseri effettivamente intelligenti, in altre parole, chiedersi se le macchine potessero pensare era una questione mal posta, e suggerì di domandarci se esse potessero invece superare un test comportamentale. Egli proponeva il problema sotto forma di gioco, il cosiddetto gioco dell’imitazione[[49]](#footnote-49), in questo modo Turing riuscì ad elaborare un test molto pratico che si proponeva di valutare quanto una macchina fosse in grado di imitare quelle peculiari caratteristiche umane che emergono durante un’interazione.

I protagonisti del test sono un giudice umano, e due interlocutori (un umano e un computer). Al giudice è ovviamente nascosta l’identità dei due interlocutori, non può vederli, ma può solo porre domande e ascoltare le risposte, in una conversazione svolta in un linguaggio naturale, la comunicazione inoltre avviene per iscritto per evitare che l’identità della macchina venga svelata dalla pronuncia. Il giudice dovrà poi stabilire, in base all’adeguatezza delle risposte nel dialogo, chi sia l’uomo e chi sia la macchina, se la persona che poneva le domande non fosse stata capace di distinguere le risposte della macchina da quelle dell'altra persona, la macchina in qualche modo si poteva considerare intelligente. Fu questa una delle prime volte in cui emerse il concetto di *intelligenza artificiale*. Turing, infatti, era dell'idea che si potessero creare macchine che fossero capaci di simulare perfettamente i processi del cervello umano. È quasi superfluo specificare che questo test non valuta la correttezza delle risposte, ma solo quanto queste siano somiglianti a quelle che potrebbe dare un essere umano. Il metodo delle domande e delle risposte permette di considerare esclusivamente le capacità intellettuali dei due interlocutori, la macchina in questione è programmata in modo da imitare le eventuali risposte che un uomo darebbe a delle domande riguardanti un qualsiasi campo della conoscenza umana. Se il giudice umano non riesce a distinguere la conversazione della macchina da quella dell’altro umano, allora si può dichiarare che la macchina ha superato il test di Turing.

Turing afferma che se una macchina è capace di sostenere adeguatamente il gioco dell’imitazione, debba allora essere ritenuta intelligente, considerando l’esito positivo del test come una condizione sufficiente per sostenerlo, ma non necessaria. Se infatti chiedessimo a un essere umano di svolgere dei calcoli alla velocità di un calcolatore, l’umano fallirebbe quasi sicuramente, per questo motivo Turing considera il test abbastanza sfavorevole per le macchine, e a tal proposito ne giudica l’esito positivo sufficiente ma non necessario.

Il gioco dell’imitazione di Turing richiama inevitabilmente alla mente una riflessione di Cartesio, riportata nella quinta parte del discorso sul metodo. Cartesio considera la possibilità di riuscire a costruire, un giorno, degli automi in grado di agire esattamente come un essere umano, imitandolo alla perfezione. Secondo il filosofo però, sarebbe possibile distinguere un automa da un essere umano perché gli automi mancherebbero di reale conoscenza, o consapevolezza, e agirebbero solo in base a delle istruzioni interne, frutto di una programmazione, a cui Cartesio si riferisce parlando di disposizione degli organi.

«E qui non mi ero fermato in particolare a mostrare che, se ci fossero delle macchine dotate degli organi e della figura esterna di una scimmia o di qualche altro animale irragionevole, non avremmo alcun mezzo per riconoscere che non sono della stessa natura di quegli animali; mentre che, se ve ne fossero di quelle fatte a somiglianza dei nostri corpi le quali imitassero le nostre azioni per quanto è moralmente possibile, avremmo sempre due mezzi sicurissimi per riconoscere che non sono dei veri uomini: il primo è che esse macchine non potrebbero mai adoperare parole o altri segni che le compongano, come noi facciamo per esprimere agli altri il nostro pensiero; poiché si può immaginare una macchina fatta in modo che pronunci delle parole, ed anche ne pronunci alcune in relazione agli atti corporei che producono qualche mutamento nei suoi organi; come, toccandola in qualche punto, che domandi che cosa le si vuol dire; toccandola in un altro, che gridi che le si fa male, e simili; ma non si otterrà mai che essa combini le parole diversamente per rispondere al significato di tutto ciò che si dirà in sua presenza , come possono fare gli uomini anche i più stupidi. Ed il secondo mezzo è che, anche se tali macchine riuscissero a fare alcunché come noi o anche meglio di noi, mancherebbero certamente in altre da cui si scoprirebbe che esse non agiscono con conoscenza ma solamente per disposizione dei loro organi.»[[50]](#footnote-50)

### **3.1.1 Le risposte di Turing a nove obiezioni**

Turing era convinto che entro la fine del ‘900 sarebbe stato possibile costruire dei calcolatori tali che un esaminatore medio non avrebbe avuto più del 70% di possibilità di riconoscere una macchina. Fin dall’inizio dell’IA, Turing aveva pensato a programmi paragonabili al cervello dei bambini, che come i bambini evolvessero e crescessero imparando, disponendo di istruzioni per fare tesoro di prove ed errori.

Alcuni chatbot costruiti tra gli anni sessanta e duemila come ELIZA, MGONZ e ALICE, hanno effettivamente indotto in errore degli esseri umani per circa cinque minuti, tuttavia nessun programma si è avvicinato alla soglia del 30 % con giudici addestrati.

Turing sosteneva che in futuro il gioco dell’imitazione sarebbe stato condotto così bene, tanto che si sarebbe potuto benissimo parlare di macchine pensanti. Nell’articolo del 1950, tenta di anticipare le eventuali obiezioni che si sarebbero potute muovere contro la sua concezione di macchina pensante (alcune delle quali sono effettivamente valide tutt’oggi), e formula altrettante risposte.

1. Obiezione religiosa: il pensiero umano è una funzione dell’anima immortale e di conseguenza le macchine, non avendo l’anima, non possono pensare.

Risposta: Turing afferma che certamente non è nostro compito creare le anime, possiamo tuttavia creare le “dimore” delle anime allo stesso modo in cui, quando procreiamo, creiamo nuovi esseri umani ma non le loro anime. «Nel tentativo di creare queste macchine non possiamo irriverentemente usurpare il Suo potere di creare le anime, non più che nella procreazione di bambini; siamo in entrambi i casi strumenti della Sua Volontà nel creare dimore per le anime che Egli crea»[[51]](#footnote-51).

1. Obiezione della “testa nella sabbia”: le conseguenze della creazione di macchine pensanti sarebbero pericolose e terrificanti, sarebbe opportuno non riuscirci mani.

Risposta: secondo Turing, queste idee sono diffuse tra gli intellettuali perché pensano che la loro superiorità derivi da una più alta intelligenza e di conseguenza vivono come una minaccia la possibilità dell’esistenza di macchine pensanti, magari con un’intelligenza superiore alla loro.

1. Obiezione matematica: questa obiezione considera il teorema di incompletezza di Gödel per dimostrare che ci sono dei limiti a quello che un computer programmato con algoritmi logici può fare. È noto, anche grazie al lavoro di Turing stesso, che ad alcune questioni matematiche non si può rispondere attraverso i sistemi formali. Adoperando metodi formali, come quello di Godel, si sono infatti verificati i limiti dei sistemi stessi. Il filosofo J. R. Lucas sosteneva che il teorema di Gödel dimostrasse che le macchine, in quanto funzionanti sulla base di sistemi formali, fossero limitate e quindi mentalmente inferiori agli umani. Anche Penrose riprese l’argomentazione gödeliana per affermare l’impossibilità per le macchine di eguagliare il cervello umano.

Risposta: è vero che ci sono dei limiti ben precisi alla possibilità di potenza delle macchine, ma non è detto che anche il cervello umano non abbia limitazioni. Ammettere che non le abbia è solo una supposizione per la quale non c’è alcuna prova.

1. Argomento della coscienza: solamente quando una macchina sarà in grado di scrivere una poesia o di comporre un concerto secondo i propri pensieri ed emozioni e non sulla base della combinazione più o meno casuale di simboli, soltanto allora si potrà dire che una macchina è equiparabile a un cervello: non deve solo aver scritto, ma avere la conoscenza di aver scritto[[52]](#footnote-52).

Risposta: per Turing, tale ragionamento portato agli estremi indurrebbe a concludere che per essere sicuri che una macchina pensi ci sarebbe la necessità di essere personalmente quella macchina e sentirsi pensare. Ma ciò varrebbe anche nei confronti degli altri uomini! Infatti, per essere sicuri che un altro essere umano all’infuori di noi pensi, bisognerebbe essere in prima persona quell’uomo, tuttavia si adotta per convenzione che tutti pensino.

1. Argomento su varie incapacità delle macchine: si basa sull’asserzione che una macchina non potrà mai fare determinate cose come «essere gentile, pieno di risorse, bello, amichevole, avere iniziativa, senso dello humor, riconoscere ciò che è giusto e sbagliato, fare errori, innamorarsi, godersi una coppa di gelato alle fragole, far sì che qualcuno la apprezzi, imparare dall’esperienza, usare le parole correttamente, essere l’oggetto del proprio pensiero, esibire una diversità di comportamento come quella di un essere umano, fare qualcosa di veramente nuovo».[[53]](#footnote-53)

Risposta: Turing risponde a questa obiezione affermando che essa derivi dall’idea generale che si aveva delle macchine del suo tempo. Durante la sua epoca infatti le macchine erano brutte, limitate nell’uso ecc., in particolare, Turing riconduce l’idea che le macchine siano prive della capacità di mettere in atto comportamenti vari semplicemente al fatto che, sempre nella sua epoca, esse non erano dotate di una grande capacità di memoria. Sarebbe dunque bastata una maggiore capacità di memoria per ovviare a questa obiezione, in quanto molte delle attività che apparentemente richiedono capacità di giudizio umana, sono riproducibili tramite l’esecuzione di algoritmi.

1. Obiezione di Lady Lovelace: la macchina analitica, come quella creata da Charles Babbage, non può creare niente di originale, ma può fare soltanto ciò che noi le ordiniamo.

Risposta: Turing afferma che sarà possibile che le macchine imparino, dedicando un paragrafo al machine learning.

1. Obiezione sulla diversità del sistema nervoso: le ricerche neurofisiologiche dimostrano che il cervello non è un computer digitale.

Risposta: Turing confida nel fatto che computer digitali potranno presto simulare comportamenti analogici.

1. Obiezione sulla prevedibilità del comportamento: secondo questa obiezione, il comportamento umano sarebbe troppo complesso per essere catturato da un semplice insieme di regole, quindi non è possibile formalizzarlo. Ogni sistema governato da regole rigide invece, come nel caso del calcolatore, è prevedibile. Da ciò deriva che la macchina non è veramente intelligente proprio in quanto il suo comportamento è fissato all’interno di un insieme di regole, mentre nessun insieme di regole può guidare la condotta umana.

In IA, l’incapacità di riuscire a catturare la varietà dei comportamenti informali in un insieme di regole logiche, si chiama problema della qualificazione.

A tal proposito si è espresso anche il filosofo Hubert Dreyfus[[54]](#footnote-54), affermando che le competenze umane sono derivabili effettivamente da alcune regole, le quali sono però da considerare solamente come *contesto olistico*. Gran parte dei processi mentali si svolgono ad un livello che non è accessibile all’introspezione conscia della mente, ma solo per questo motivo non si può dedurne che i *processi* mentali non esistano.

Risposta: anche gli esseri umani, seppur in modo non propriamente conscio, sottostanno a delle leggi di natura applicate al corpo, secondo Turing quindi, anche gli esseri umani sono in qualche modo macchine, anche se non digitali.

1. Obiezione riguardante la percezione extrasensoriale: le percezioni extrasensoriali quali chiaroveggenza, telepatia, telecinesi, precognizione, erano prese in seria considerazione all’epoca di Turing.

Risposta: Turing è evidentemente disposto a credere che la differenza ultima tra esseri umani e macchine risulterà essere proprio la percezione extrasensoriale. Egli era infatti convinto che le prove a favore della telepatia fossero molto forti.

## 3.2 Il funzionalismo computazionale

Alan Turing aveva dimostrato che il suo dispositivo meccanico di calcolo era in grado di risolvere qualsiasi problema esprimibile in istruzioni scritte in linguaggio logico-matematico, svolgendo un processo di elaborazione sintattica dell’informazione. Turing era anche dell'idea che si potessero creare macchine in grado di simulare perfettamente i processi del cervello umano, ma per affermare questo bisogna stabilire *come* funziona il cervello umano.

Tra gli argomenti centrali della filosofia della mente, vi è certamente quello riguardante la natura dei fenomeni mentali e come questi fenomeni mentali si inseriscano nella struttura causale della realtà.

Sulla scia dell’innovazione da lui introdotta, Turing iniziò a maturare la convinzione che ci potesse essere un punto di contatto tra il suo sistema artificiale e le modalità di funzionamento del pensiero umano, il quale poteva essere considerato come un’elaborazione sintattica dell’informazione processato da una macchina deterministica: il cervello.

Se vogliamo dire che un determinato programma ragiona come un essere umano, dobbiamo prima determinare come pensiamo, entrare cioè dentro i meccanismi del cervello stesso[[55]](#footnote-55), e solo una volta che abbiamo formulato una teoria della mente sufficientemente precisa allora diventa possibile esprimerla sotto forma di un programma per computer[[56]](#footnote-56), ma ciò sarà possibile solo a patto di considerare tutte le operazioni del cervello come funzioni assolutamente computabili.

Hilary Putnam dichiarò che, analogamente ad una macchina di Turing, anche noi umani siamo automi a stati finiti, quello che intendeva dire è che le nostre funzioni cognitive possono essere calcolate da una macchina di Turing. Secondo la teoria computazionale della mente, tutte le funzioni cognitive possono essere calcolate da algoritmi. L’algoritmo è un procedimento che risolve un determinato problema attraverso un numero finito di passi elementari e per calcolare una funzione, bisogna eseguire uno o più algoritmi. La teoria computazionale della mente, riguarda gli algoritmi con cui calcoliamo le funzioni cognitive: analogamente alle capacità computazionali di una macchina, l'esercizio delle abilità mentali consisterebbe nel calcolo di queste funzioni.

La teoria funzionalista si pone l’obiettivo di individuare il meccanismo computazionale della mente e di conseguenza progettare l’architettura funzionale di un sistema informatico che sia in grado di riprodurre le nostre capacità mentali. Questa teoria parte dal presupposto che il cervello umano funzioni come un computer, il quale opera generando relazioni causali tra eventi e stati mentali.

Essenzialmente, la teoria afferma che: «Tutto il pensiero umano è riconducibile ad attività di tipo algoritmico, cioè a trasformazioni di simboli in base a regole formali esplicitamente formulabili ed eseguibili in modo puramente meccanico»[[57]](#footnote-57). Stando a questa definizione quindi, potremmo definire uno stato mentale come una qualsiasi condizione causale intermedia tra input e output. Di conseguenza potremmo affermare che secondo i princìpi del funzionalismo, due sistemi qualsiasi con processi causali isomorfi avrebbero gli stessi stati mentali, di conseguenza anche un computer potrebbe avere gli stati mentali di una persona. Gli stati mentali umani sono quindi duplicabili da un programma con la stessa struttura funzionale di un cervello e con lo stesso comportamento di input-output. C’è inoltre da considerare che medesimi programmi funzionano, esattamente nello stesso modo, su macchine diverse, e ciò proverebbe il fatto che se lo stato mentale è un insieme di funzioni esattamente come un programma, allora lo stato mentale è indipendente dal suo substrato materiale.

L’interrogativo fondamentali da porsi a questo punto è se davvero il cervello umano, o meglio se la “macchina cervello” deve tutte le sue straordinarie capacità ad un funzionamento basato su algoritmi e istruzioni computabili, se la risposta a questa domanda è affermativa, se cioè è possibile spiegare l’intelligenza umana come risultato di una complessa elaborazione e manipolazione di dati alla stregua di quanto accade nei sistemi formali, e quindi se i processi cerebrali sono esprimibili sotto forma di algoritmo, allora potremmo anche dire che è possibile, in linea di principio, trasferire la nostra intelligenza alle macchine. Alan Turing era dell’idea che, se un giorno si fosse riusciti a riprodurre una macchina in grado di riprodurre i complicati circuiti elettro-chimici del cervello umano, allora sarebbe stato possibile trasferire a questa macchina un’intelligenza di tipo analogo.

### **3.2.1 Esperimento della protesi cerebrale**

Il seguente è un esperimento teorico sul pensiero, che affronta la questione della possibilità per oggetti fisici differenti dai neuroni umani di avere degli stati mentali.

A metà degli anni ’70, Clark Glymour introdusse il cosiddetto esperimento della protesi cerebrale, formulandone la definizione in questo modo: supponiamo che la scienza neurofisiologica si sia sviluppata a tal punto da comprendere perfettamente il comportamento di input-output e la connettività di tutti i neuroni del cervello umano, insomma che sia stato scoperto il modo per formalizzare totalmente i meccanismi cerebrali. Supponiamo poi che si possano costruire dei microscopici dispositivi artificiali, elettronici, in grado di imitare il comportamento dei neuroni, e che sia possibile collegarli direttamente al tessuto neurale. Poi supponiamo ancora che con una particolare tecnica chirurgica si sia in grado di sostituire gradatamente ogni singolo neurone con i dispositivi elettronici, senza mai interrompere il funzionamento del cervello nel suo insieme, arrivando a sostituire ogni neurone. Secondo la definizione dell’esperimento, il comportamento esterno del soggetto dovrebbe rimanere immutato rispetto al comportamento che l’uomo avrebbe se l’operazione non fosse stata eseguita. Ora, siccome l’assenza o la presenza di coscienza in una persona non può essere facilmente determinata da una terza persona, almeno il soggetto dell’esperimento dovrebbe essere in grado di notare eventuali cambiamenti nella propria esperienza interna conscia.

Patricia Churchland, sostenitrice della teoria funzionalista, mette in evidenza il fatto che le argomentazioni funzionaliste che operano a livello de singolo neurone, possono anche operare a livelli superiori. Ciò significa che se in seguito a questo esperimento, dovremmo constatare che la coscienza è conservata.

Secondo la teoria funzionalista, dal momento che le funzioni cerebrali sono state perfettamente riprodotte dai meccanismi elettronici, anche la coscienza, che emerge da tali funzioni, dovrebbe rimanere inalterata. In questo caso dovremmo ammettere che i meccanismi causali della coscienza stanno operando ancora nella versione elettronica, e di conseguenza spiegare le manifestazioni di coscienza prodotte dal cervello elettronico facendo riferimento unicamente alle proprietà funzionali dei neuroni.

## 3.3 Il naturalismo biologico di J. Searle

Nel test di Turing è insita la tentazione di pensare che se qualcosa si comporta come se avesse certi processi mentali allora debba avere davvero questi processi mentali. Questa concezione fa parte dell’erroneo assunto del comportamentismo che, per essere scientifica, la psicologia debba limitarsi a studiare il comportamento esterno osservabile. Ritorniamo dunque alla domanda fondamentale posta in questo capitolo: può una macchina pensare? Se consideriamo la questione dal punto di vista del naturalismo biologico e per macchina si intende un sistema fisico capace di compiere certe funzioni, allora gli umani potrebbero essere considerati come un particolare tipo di macchina biologica e non ci resterebbe che ammettere che si, queste macchine sono in grado di pensare. Il filosofo J. Searle non rifiuta l’idea che un giorno le macchine possano pensare, per il semplice fatto che considera anche l’essere umano come una macchina, una macchina biologica. Non esclude quindi la possibilità che un giorno si possa creare una macchina pensante e artificiale, con materiale che non sia organico. Egli sostiene però che negli ultimi decenni, il quesito riguardante la possibilità che una macchina possa pensare ha subìto un’interpretazione diversa, sotto l’influenza della tendenza funzionalista la domanda primaria è stata addirittura sostituita da un altro quesito, diventando la seguente: una macchina può pensare semplicemente in virtù del fatto che esegue un programma di calcolatore? il programma è di per sé una componente del pensiero? Ciò che Searle vuole confutare quindi, è la teoria funzionalista, e non l’idea che possano essere create delle macchine pensanti.

Searle, in sostanza non ritiene che la mente possa essere ridotta ad un programma, cioè ad un insieme di funzioni formalizzabili.

Al contrario della teoria funzionalista, la teoria del naturalismo biologico afferma che gli stati mentali sono caratteristiche emergenti causate da processi neurologici di basso livello all’interno dei neuroni, e che ciò che è rilevante nella produzione della coscienza sono le proprietà dei neuroni stessi, seppur non specificate. Secondo questa teoria gli stati mentali non sono riproducibili da un programma che riproduca la struttura funzionale del cervello, e due stati mentali non potranno essere considerati dei duplicati semplicemente perché hanno lo stesso comportamento di input-output (la teoria però non spiega perché e come facciano i neuroni ad avere questo potere causale)[[58]](#footnote-58).

Il cervello, essendo un organo biologico specifico *causa* eventi mentali grazie a processi neurobiologici specifici, le sue proprietà biochimiche gli consentono di causare la coscienza e altri tipi di fenomeni mentali. Le simulazioni al calcolatore dei processi cerebrali forniscono modelli degli aspetti formali di questi processi, ma la simulazione non va confusa con la riproduzione, Searle rende bene l’idea con queste parole: «Nessuno pensa che una simulazione al calcolatore della digestione possa realmente digerire qualcosa, ma quando si ha a che fare con i processi cognitivi si è disposti a credere in miracoli del genere perché non ci si rende conto che la mente è un fenomeno biologico al pari della digestione.»[[59]](#footnote-59) La simulazione di un processo cognitivo infatti non produce gli stessi effetti della neurobiologia di quel processo cognitivo.

Secondo il naturalismo biologico, è il cervello che causa la mente, e qualunque altro sistema in grado di causare una mente dovrebbe avere poteri causali almeno equivalenti a quelli del cervello. Qualunque sistema artificiale in grado di produrre fenomeni mentali, ossia qualunque cervello artificiale, dovrebbe essere in grado di riprodurre gli stessi poteri causali specifici del cervello umano e non potrebbe farlo soltanto svolgendo un programma formale, perché il modo in cui il cervello umano produce i fenomeni mentali non può ridursi solamente allo svolgimento di un programma al calcolatore.

In altre parole, «poiché tutto ciò che può essere simulato per via computazionale può essere descritto come un calcolatore e poiché il nostro cervello può a certi livelli essere simulato, ne segue ovviamente che il nostro cervello è un calcolatore ed esso può certo pensare. Ma dal fatto che un sistema possa essere simulato per mezzo della manipolazione di simboli e dal fatto che esso sia pensante non ne segue che pensare equivalga all’esecuzione di manipolazioni formali su simboli»[[60]](#footnote-60)

Secondo la teoria del naturalismo biologico, in un caso come quello descritto dall’esperimento della protesi cerebrale, la coscienza svanirebbe, in quanto si tratterebbe semplicemente di *simulare* i processi reali del nostro cervello. Ma cos’è la coscienza per Searle? Egli parte dall’osservazione che se essa esiste, sicuramente non sembra essere un qualcosa di fisico. Osserva che per tale motivo molti filosofi hanno abbracciato la versione del dualismo, ossia l’idea che nell’universo esistano due fenomeni differenti e complementari: il mentale e il fisico, è il caso ancora una volta di prendere in causa Cartesio, il quale aveva considerato l’uomo come insieme di *res cogitans* e *res extensa*. Altri autori invece, soprattutto nella filosofia contemporanea, hanno optato per il materialismo, eliminando quindi la coscienza, o semplicemente riducendola a qualcos’altro, come nel caso di Daniel Dennet.

John Searle, da parte sua, rifiuta il dualismo, il quale operando una netta distinzione tra fisico e mentale, non riesce a rendere intelligibile la relazione tra i due, ma rifiuta anche il materialismo, e di conseguenza la visione funzionalista che riduce la coscienza ad uno stato mentale descrivibile in termini puramente fisici. Egli afferma che questa tendenza contemporanea verso il riduzionismo e il materialismo derivi dall’errore fondamentale di supporre che, se la coscienza abbia una propria esistenza reale, allora dovremmo necessariamente accettare in qualche modo il dualismo e rinunciare ad una visione scientifica del mondo. Searle offre un’interpretazione nuova di coscienza, considerandola alla stregua di ogni altro fenomeno naturale, biologico, esattamente come la digestione o la fotosintesi. Secondo Searle, abbiamo difficoltà a cogliere il carattere naturale della coscienza proprio a causa della nostra tradizione filosofica che ha fatto del mentale e del fisico due categorie separate che si escludono reciprocamente. La coscienza, tuttavia, non rientra in nessuna di queste due categorie tradizionali, e la via d’uscita starebbe proprio nel rifiutare sia la visione dualista che il più diffuso materialismo. Searle pone la coscienza al centro della questione circa la comprensione della nostra esistenza autentica di esseri umani, in quanto è la condizione che rende possibile che una qualsiasi cosa abbia significato per ognuno di noi, soltanto per gli agenti coscienti, infatti, si pone la questione dell’importanza o del significato di qualcosa, quindi della dimensione semantica. Secondo il senso comune del termine, la parola *coscienza* si riferisce a tutti quegli stati di consapevolezza e sensibilità che ci caratterizzano durante il nostro stato di veglia. Secondo il filosofo, «essa è causata da microprocessi di livello inferiore nel cervello ed è una caratteristica del cervello a macrolivelli superiori»[[61]](#footnote-61), di conseguenza noi saremo in grado di comprendere realmente la coscienza solo quando capiremo come funziona realmente il cervello in termini biologici. La coscienza è considerata essa stessa una caratteristica del cervello, Searle la definisce come proprietà emergente, in quanto emerge dalle attività neuronali. A questo punto ci si potrebbe chiedere per quale motivo, questa proprietà non possa emergere anche da un sistema artificiale di neuroni. La risposta negativa è giustificata dal fatto che tale proprietà è spiegata causalmente dagli elementi del sistema, ma non è una proprietà di qualsiasi elemento individuale e non può essere spiegata semplicemente come la somma delle proprietà di quegli elementi!

Uno degli errori più diffusi nella considerazione della coscienza, deriva da una concezione errata di causazione. Se sono i processi cerebrali a causare la coscienza, allora ciò potrebbe indurre nell’errore di credere che ci siano effettivamente due cose separate: i processi cerebrali come causa e la coscienza come effetto. In realtà, secondo Searle, i processi di livello inferiore del cervello causano gli stati di coscienza presente, ma questo stato non è un’entità separata dal cervello stesso! è semplicemente una caratteristica emergente del cervello al tempo presente. Un ulteriore difficoltà deriva dal fatto che i processi cerebrali, pur essendo oggettivi e osservabili, provocano gli stati interiori che sono degli stati particolari, privati, ai quali non hanno accesso altri oltre al soggetto interessato. Questi fenomeni privati, soggettivi e qualitativi sono detti *qualia*. Anche in questo caso, si potrebbe pensare che i qualia siano qualcosa di separato dalla coscienza, ma Searle asserisce ancora una volta che non ci sono due fenomeni separati: c’è solo la coscienza che è costituita da una serie di stati qualitativi.

## 3.4 IA forte e IA debole: la stanza cinese

In seguito all’affermazione della prospettiva funzionalista, si è diffusa l’idea che il cervello possa essere considerato alla stregua di una macchina, e che la coscienza sia paragonabile semplicemente ad un programma eseguito dalla macchina-cervello. Secondo questa visione, la mente starebbe al cervello esattamente come in informatica il software starebbe all’hardware. Ma se la nostra mente è un programma, è possibile ricrearla artificialmente? I sostenitori della teoria computazionale della mente direbbero ovviamente di sì. La visione secondo la quale la mente sarebbe esattamente un programma per computer (e in quanto tale è ritenuta possibile la creazione di una vera mente artificiale in grado di fare ciò che essa *realmente* pensa), è detta Intelligenza Artificiale forte. Secondo l’IA forte il pensiero umano, integralmente trasferibile in una macchina, non è altro che la manipolazione di simboli formali. Questa posizione viene spesso riassunta con la frase “la mente sta al cervello come il programma sta al calcolatore”.

Al contrario, la visione secondo la quale il computer è solo uno strumento in grado di svolgere *simulazioni* della mente e del pensiero, è detta Intelligenza Artificiale debole.

Searle confuta la teoria dell’IA forte, in quanto ritiene che un computer, anche il più avanzato, sia semplicemente un dispositivo che manipola simboli formali, e come ben sappiamo dalla nostra esperienza personale, la mente è qualcosa che va oltre la semplice manipolazione di simboli poiché per noi è fondamentale la questione del significato, ovvero la semantica. La computazione è costituita da una serie di operazioni puramente sintattiche, e Searle reputa inverosimile l’idea che sia possibile ridurre la mente ad un programma per computer, seppur molto sofisticato, perché i simboli formali di un programma non sono sufficienti a garantire la presenza del contenuto semantico che si trova nelle menti umane. I simboli formali infatti non posseggono una semantica intrinseca.

Searle sostiene che l’IA forte si discosta dalle altre teorie della mente sotto almeno due aspetti: può essere espressa in modo chiaro ed è soprattutto soggetta ad una confutazione semplice e decisiva che effettua per mezzo di un esperimento, grazie al quale riesce a mettere in luce i punti deboli della teoria computazionale della mente, rendendo al contrario evidente quanto sia fondamentale la questione della semantica.

«Immaginate di eseguire un programma per rispondere a delle domande in una lingua che non conoscete. Io non capisco il cinese, quindi immagino di essere chiuso a chiave in una stanza e di avere a disposizione molte scatole di simboli in cinese (il database), di ricevere alcuni simboli in cinese (le domande in cinese) che mi vengono passati dall’esterno, e di poter consultare un testo con delle regole (il programma) per sapere che cosa fare. Proseguo compiendo alcune operazioni sui simboli, in accordo con le regole (ovvero, seguo dei passaggi del programma), e fornisco come risposta, a coloro che stanno fuori dalla stanza, una serie di simboli (le risposte alle domande). Sebbene sia io stesso il computer che implementa un programma per rispondere a domande in cinese, ugualmente non capisco una parola di cinese. E questo è il punto: se io non riesco a capire il cinese solamente sulla base dell’implementazione di un programma per computer per comprendere il cinese, allora non lo può fare nemmeno nessun altro computer digitale solamente su quella base, perché nessun computer digitale possiede qualcosa che io non ho».[[62]](#footnote-62)

Con questo argomento, Searle vuole mettere in evidenza che i programmi, analogamente al protagonista dell’esperimento, agiscono solo ed esclusivamente sulla base di manipolazioni sintattiche.

Dall’argomento della stanza cinese, Searle ricava degli assunti fondamentali che egli chiama assiomi. Il primo assioma dichiara che i programmi sono definiti in termini di manipolazioni di simboli e i simboli sono enti puramente formali, cioè sintattici: un computer elabora informazione codificando in primo luogo nel simbolismo che esso usa e poi manipolando i simboli secondo un insieme di regole enunciate con precisione. Queste regole costituiscono il programma. I simboli e programmi inoltre, sono nozioni puramente astratte, ciò significa che non esistono proprietà fisiche essenziali che li definiscano: i simboli 0 e 1, per loro natura, non hanno alcuna proprietà fisica essenziale e causale. Searle sottolinea questo punto perché sente forte la tentazione dei suoi contemporanei di identificare i calcolatori con qualche tecnologia particolare, per esempio con la microelettrica del silicio, oppure la tendenza a pensare che le proprietà e i problemi di cui si parla riguardino la fisica dei chip di silicio, o ancora che il semplice processo di calcolo identifichi un qualche fenomeno fisico che potrebbe avere poteri causali ancora sconosciuti così come hanno proprietà fisiche causali i fenomeni fisici veri e propri, come la radiazione elettromagnetica e gli atomi di idrogeno. La seconda osservazione fondamentale che si può fare riguardo alla natura della computazione digitale, è che sui simboli si opera senza significato. I simboli del programma possono rappresentare qualunque cosa il programmatore o l’utente desideri, in questo senso il programma ha una sintassi ma non una semantica. Da ciò deriva che il possesso dei soli simboli, della sola sintassi, non è sufficiente per possedere la semantica. Le semplici manipolazioni dei simboli non bastano per garantire la conoscenza del loro significato, la sintassi infatti di per sé non è condizione essenziale, né sufficiente, per la determinazione della semantica, perché gli elementi formali non hanno un contenuto o un significato intrinseco.

Da queste premesse segue che i programmi non sono condizione essenziale né sufficiente perché sia data una mente, e con ciò Searle dimostra che l’IA forte è una posizione insostenibile, quindi è falsa.

Con la metafora della stanza cinese Searle ci vuole dimostrare che la sola manipolazione di simboli non basta di per sé a garantire l’intelligenza, la percezione, la comprensione e così via. E poiché i calcolatori sono per natura dei dispositivi per operare sui simboli, la semplice operazione di far girare il programma non è garanzia sufficiente di attività cognitive.

Seppure il protagonista abbia fornito tutte le risposte giuste, in un cinese perfetto, è rimasto all’oscuro del significato sia delle domande che delle risposte da lui stesso fornite. Il comportamento adeguato di per sé non è condizione necessaria per il pensiero: questo è il principio su cui si basa l’argomento della stanza cinese. Anche Dreyfus, come Searle, si pone il problema del significato:

«anche se il mondo viene inserito nel computer in termini di bit indipendenti dal punto di vista logico, ciò non significa che si può affermare aprioristicamente che il mondo può essere in seguito ricomposto. A sostegno di questa considerazione vi è una teoria generale dell’esperienza umana intesa nel senso di “trovarsi già in una situazione” dove i fatti sono già stati interpretati. Questa teoria suggerisce anche che una data situazione di un essere umano dipende dai suoi obiettivi, i quali, a loro volta sono in funzione del corpo e dei suoi vari bisogni, e che questi bisogni non sono stabiliti una volta per sempre, ma vengono interpretati e determinati attraverso l’acculturazione e attraverso le modifiche nel modo di interpretarsi degli esseri umani. Un computer è in grado di lavorare solo in base a fatti, mentre l’uomo, fonte dei fatti stessi, non è un fatto insieme ai fatti, ma un essere che crea se stesso e l’universo dei fatti proprio attraverso la sua vita nel mondo. Non esistono fatti dotati di significato intrinseco, e non esistono forme stabili di vita che saremo un giorno in grado di programmare».[[63]](#footnote-63)

La questione della coscienza dunque, riguarda innegabilmente la presenza di una dimensione semantica della nostra mente, semantica che è tutt’altro dalla semplice sintassi corretta. Da ciò potremmo quindi concludere che i programmi non sono delle menti, in quanto la mente umana possiede contenuti semantici oltre che simboli formali, e la manipolazione di simboli formali non garantisce la presenza di contenuto semantico. Secondo Searle, non ha importanza quanto il sistema riesca ad imitare il comportamento umano, né quanto complessa possa essere la manipolazione di simboli, poiché si tratterebbe in ogni caso solo ed esclusivamente di simulazione (IA debole), e siccome non è possibile paragonare una mente ad un programma, di conseguenza non sarà possibile parlare di coscienza a proposito di macchine.

1. John McCarthy, Marvin L. Minsky, Nathaniel Rochster, Claude E. Shannon, *A proposal for Dartmouth Summer Research Project on Artificial Intelligence*, ripubblicato in “AI Magazine”, vol.27 n. 4, p.12. [↑](#footnote-ref-1)
2. Margaret Boden, *Artificial Intelligence and natural Man*, The MIT Press; second edition (March 23, 1987) [↑](#footnote-ref-2)
3. Stuart Russell, Peter Norvig, *Intelligenza Artificiale, un approccio moderno*, vol. 1, 2° edizione, Pearson Education Italia S.r.l., Milano 2005 p. 588 [↑](#footnote-ref-3)
4. Stuart Russell, Peter Norvig, *Intelligenza Artificiale, un approccio moderno*, vol. 1, 2° edizione, Pearson Education Italia S.r.l., Milano 2005 [↑](#footnote-ref-4)
5. Cfr. cap.3 [↑](#footnote-ref-5)
6. A tal proposito ritengo opportuno tenere in considerazione i più moderni sistemi sviluppati nell’ambito delle Reti Neurali, i quali, a differenza di IA basata su sistemi inferenziali in cui è nota la rappresentazione delle conoscenze all’interno di un dominio e quindi anche le regole di inferenza, accumulano informazione sui pesi della rete e non rendono conto dei passaggi logici che utilizzano. A differenza del calcolatore tradizionale di von Neumann, che lavora passo passo, questi sistemi hanno molte unità di elaborazione che funzionano in parallelo e interagiscono fra loro secondo regole che si ispirano alla neurobiologia. [↑](#footnote-ref-6)
7. Platone, Eutifrone, 5, c-d [↑](#footnote-ref-7)
8. Il principio logico del terzo escluso che afferma che due proposizioni formanti come ad esempio A e -A devono avere valore di verità opposto, cioè non esiste una terza possibilità (Tertium non datur). In altre parole, non è possibile che due proposizioni contraddittorie siano entrambe vere o non vere, in quanto esso afferma che il valore di verità di una proposizione è sempre opposto a quello della proposizione contraddittoria. [↑](#footnote-ref-8)
9. Lorenzo Pinna, *Intelligenza Artificiale*, Edizioni Cento Autori, Villaricca 2018 [↑](#footnote-ref-9)
10. Thomason, Richmond, "Logic and Artificial Intelligence", The Stanford Encyclopedia of Philosophy (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/logic-ai/>. [↑](#footnote-ref-10)
11. Andrea Cusimano, *Storia del pensiero occidentale*, ISBN 978-1-291-59170-5 p. 149 [↑](#footnote-ref-11)
12. Silvio Hénin, La calcolatrice di Leonardo, Mondo digitale, ottobre 2018

    http://mondodigitale.aicanet.net/2018-5/Rubriche/01\_MD78\_Henin\_IlCalcolatoreDiLeonardo.pdf [↑](#footnote-ref-12)
13. B. Pascal, Oeuvres complètes, cit., pp. 189-190. Trad. it. La macchina aritmetica di Blaise Pascal, a cura di M. Sangoi e P. Graziani, Isonomia 2005, p. 6. [↑](#footnote-ref-13)
14. Th. Hobbes, Leviatano, I, 5, Utet. Torino 1855, p. 76 [↑](#footnote-ref-14)
15. Ludovico Geymonat, Dizionario dei termini filosofici, p. 77, allegato a Immagini dell'Uomo, Garzanti, 1989. [↑](#footnote-ref-15)
16. René Descartes, Regole per la guida dell’intelligenza, titolo originale Regulae ad diretione ingenii, Fabbri Editori, Milano 2000, pp. 163-165 [↑](#footnote-ref-16)
17. Ibid. p. 179 [↑](#footnote-ref-17)
18. Ibid. p. 40 [↑](#footnote-ref-18)
19. Ibid. p. 16 [↑](#footnote-ref-19)
20. Ibid p. 381 [↑](#footnote-ref-20)
21. Goffredo Guglielmo Leibniz, *Scritti di logica, Sulla sintesi e sull’analisi universale, ossia sull’arte dello scoprire e del giudicare*, a cura di Francesco Barone, Zanichelli editore, Bologna 1968 pp. 216-217 [↑](#footnote-ref-21)
22. Si tratta dei telai da tessitura della seta utilizzati a partire dal XVIII sec. nel quale le diverse operazioni venivano programmate prima attraverso l’utilizzo di nastri e cilindri e successivamente attraverso l’utilizzo di schede perforate. Ciò consentiva di cambiare trama e disegni con pochissima fatica. [↑](#footnote-ref-22)
23. Martin Davis, *Il calcolatore universale, da Leibniz a Turing*, tr. it. Gianni Rigamonti, Adelphi edizioni, Milano 2003 [↑](#footnote-ref-23)
24. Lettera a L’Hôpital, 28 aprile 1693. [↑](#footnote-ref-24)
25. È credenza comune che la notazione binaria, che è alla base dell’odierna tecnologia computazionale, sia stata inventata da Leibniz. In realtà lo studioso tedesco prese spunto da un sistema già utilizzato e descritto dagli antichi cinesi nel libro dei I-Ching ben 5000 anni prima. [↑](#footnote-ref-25)
26. Martin Davis, *Il calcolatore universale*, da Leibniz a Turing, tr. it. Gianni Rigamonti, Adelphi edizioni, Milano 2003 [↑](#footnote-ref-26)
27. George Boole, *L’analisi matematica della logica*, Bollati Boringhieri editore, Torino 1993 pp-3-4 [↑](#footnote-ref-27)
28. Andrea Pedeferri, *George Boole*, APhEx, Portale italiano di filosofia analitica, n. 2 giugno 2010 http://www.aphex.it/public/file/Content20141210\_08.APhEx2,2010ProfiliBoolePedeferri.pdf [↑](#footnote-ref-28)
29. Francesca Boccuni, *Gottlob Frege*, APhEx, Portale italiano di filosofia analitica, n. 3 gennaio 2011 [↑](#footnote-ref-29)
30. Frege, *Begriffsschrift*, 1879, trad. it. p. 104 [↑](#footnote-ref-30)
31. Martin Davis, *Il calcolatore universale*, da Leibniz a Turing, tr. it. Gianni Rigamonti, Adelphi edizioni, Milano 2003 [↑](#footnote-ref-31)
32. Questo pensiero si trova già in Gottfried Leibniz che cercava una characteristica universalis, una scienza universale, da cui potessero essere dedotte tutte le altre scienze come istanze specifiche. [↑](#footnote-ref-32)
33. http://www.treccani.it/enciclopedia/logicismo\_%28Enciclopedia-della-Matematica%29/ [↑](#footnote-ref-33)
34. Bertrad Russel- Alfred North Whitehead*, Principia Mathematica*, Cambridge University press, 1910, vol. I, Preface. [↑](#footnote-ref-34)
35. Linsky, Bernard and Irvine, Andrew David, "Principia Mathematica", The Stanford Encyclopedia of Philosophy (Fall 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2019/entries/principia-mathematica/>. [↑](#footnote-ref-35)
36. https://www.riflessioni.it/dizionario\_filosofico/formalismo.htm [↑](#footnote-ref-36)
37. Ignazio Licata, La logica aperta della mente, Codice edizioni, Torino 2008, pg 84 [↑](#footnote-ref-37)
38. Martin Davis, Il calcolatore universale, da Leibniz a Turing, tr. it. Gianni Rigamonti, Adelphi edizioni, Milano 2003 p. 117 [↑](#footnote-ref-38)
39. La seconda rivoluzione scientifica: matematica e logica. L'intuizionismo di Brouwer

    di Anne L. Troelstra - Storia della Scienza (2004), http://www.treccani.it/enciclopedia/la-seconda-rivoluzione-scientifica-matematica-e-logica-l-intuizionismo-di-brouwer\_%28Storia-della-Scienza%29/ [↑](#footnote-ref-39)
40. Immerman, Neil, "Computability and Complexity", The Stanford Encyclopedia of Philosophy (Winter 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2018/entries/computability/>. [↑](#footnote-ref-40)
41. https://www.enzopennetta.it/2018/01/godel-e-lintelligenza-artificiale/ [↑](#footnote-ref-41)
42. Il teorema di Gödel sarà ripreso da alcuni pensatori, come ad esempio Penrose, a difesa delle teorie contrarie a quelle che sostengono la possibilità della creazione di un AI forte. [↑](#footnote-ref-42)
43. Raatikainen, Panu, "Gödel's Incompleteness Theorems", The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2018/entries/goedel-incompleteness/>. [↑](#footnote-ref-43)
44. Prendiamo come esempio pratico il paradosso di Epimenide che, essendo Cretese, affermò: "Tutti i Cretesi sono bugiardi". È un enunciato autoreferenziale: se Epimenide dice il vero contraddice l'affermazione che diche che tutti i Cretesi sono mentitori, se invece dice il falso allora non è vero che tutti i Cretesi sono bugiardi. Il teorema di Gödel vuole dire che per  
    quanto si affinino gli strumenti della matematica rimarrà sempre una proposizione indecidibile nella realtà. [↑](#footnote-ref-44)
45. A proposito di ciò, ritengo opportuno citare una frase di Aristotele che negli Analitici II affermava: «non esiste alcun genere di conoscenza superiore alla dimostrazione se non l’intuizione» [↑](#footnote-ref-45)
46. Martin Davis, *Il calcolatore universale*, da Leibniz a Turing, tr. it. Gianni Rigamonti, Adelphi edizioni, Milano 2003 [↑](#footnote-ref-46)
47. A. M. Turing, *Computing Machinery and Intelligence*, Mind 49: 433-460, 1950 [↑](#footnote-ref-47)
48. Ci tengo a precisare che questa differenziazione del sapere in vari settori è puramente formale perché ritengo puramente formali i limiti e le demarcazioni di confine tra le varie discipline. [↑](#footnote-ref-48)
49. The imitation game [↑](#footnote-ref-49)
50. Renato Descartes, *Il discorso sul metodo*, trad., intro e note di Guido de Giuli, G. B. Paravia & C., Torino 1945, parte V, pp. 55-56 [↑](#footnote-ref-50)
51. [↑](#footnote-ref-51)
52. Turing cita espressamente un discorso del Professor Geoffrey Jefferson del 1949 [↑](#footnote-ref-52)
53. A. M. Turing, Computing Machinery and Intelligence, Mind 49: 433-460, 1950 [↑](#footnote-ref-53)
54. Hubert L. Dreyfus, *Che cosa non possono fare i computer, i limiti dell’intelligenza artificiale*, Armando Editore, Roma 1988 [↑](#footnote-ref-54)
55. Il campo interdisciplinare delle scienze cognitive unisce modelli per computer sviluppati dall’intelligenza artificiale e tecniche di sperimentazione psicologica nel tentativo di costruire teorie precise e verificabili sul funzionamento della mente umana. [↑](#footnote-ref-55)
56. Stuart Russel, Peter Norvig, Intelligenza Artificiale, un approccio moderno, Pearson Education Italia, Milano 2005, ISBN 88-7192-228-X [↑](#footnote-ref-56)
57. Marco Giunti, Roberto Giuntini, Macchine, calcolo e pensiero, https://www.academia.edu/27149871/Macchine\_calcolo\_e\_pensiero [↑](#footnote-ref-57)
58. Stuart Russel, Peter Norvig, Intelligenza Artificiale, un approccio moderno, Pearson Education Italia, Milano 2005, ISBN 88-7192-228-X [↑](#footnote-ref-58)
59. John R. Searle, *La mente è un programma?, Le scienze quaderni* n 66, giugno 1992 [↑](#footnote-ref-59)
60. John R. Searle, La mente è un programma?, Le scienze quaderni n 66, giugno 1992 [↑](#footnote-ref-60)
61. John R. Searle, Il mistero della coscienza, Raffaello Cortina editore, Milano 1998, p. XIII [↑](#footnote-ref-61)
62. John R. Searle, Il mistero della coscienza, Raffaello Cortina editore, Milano 1998, p.8 [↑](#footnote-ref-62)
63. Hubert L. Dreyfus, *Che cosa non possono fare i computer, i limiti dell’intelligenza artificiale*, Armando Editore, Roma 1988Ignazio Licata, La logica aperta della mente, Codice edizioni, Torino 2008 p.398 [↑](#footnote-ref-63)