

Giovanni Landi

INTELLIGENZA ARTIFICIALE  
COME FILOSOFIA

“Orizzonti”

55



TANGRAM  
EDIZIONI SCIENTIFICHE  
TRENTO

Giovanni Landi, *Intelligenza Artificiale come Filosofia*  
Copyright © 2020 Tangram Edizioni Scientifiche  
Gruppo Editoriale Tangram Srl  
Via dei Casai, 6 – 38123 Trento  
[www.edizioni-tangram.it](http://www.edizioni-tangram.it)  
[info@edizioni-tangram.it](mailto:info@edizioni-tangram.it)

Collana “Orizzonti” – NIC 55

Prima edizione: gennaio 2020, *Printed in EU*

ISBN 978-88-6458-196-5

In copertina: Foto di Brett Jordan, Unsplash.com



Guida alla lettura	9
Prefazione	11
Introduzione	17
1 Filosofia della storia nella (e della) Intelligenza Artificiale	23
Elementi di filosofia nella storia dell'Intelligenza Artificiale	26
In cammino verso l'Intelligenza Artificiale	34
2 Storia della filosofia nella (e della) Intelligenza Artificiale	49
I teologi razionali	51
I logici	71
3 Logica, matematica, filosofia	85
Conclusione	105
Dialettica digitale	110
Appendice etica	133
Intelligenza Artificiale: l'empatia e il meccanismo	143
L'empatia in filosofia	147
L'empatia in psicologia	153
I neuroni specchio, ossia l'empatia "incarnata"	154
Tentativi di empatia artificiale	163
Macchina e meccanismo	179
Bibliografia	185

INTELLIGENZA ARTIFICIALE  
COME FILOSOFIA

## GUIDA ALLA LETTURA

“Intelligenza Artificiale come Filosofia”: letto così, il titolo suggerisce un rapporto di identità tra le due discipline che si tratterà di sottolineare e giustificare. È una lettura corretta.

Ma c'è un secondo senso, più sottile, che il titolo porta con sé: la necessità, di fronte a questa rivoluzione sociale, economica, tecnica e politica che chiamiamo Intelligenza Artificiale, di mantenere un contegno non ostile ma senza rassegnazione, con una ferma volontà di capire. L'Intelligenza Artificiale, insomma, va presa “con filosofia” per apprezzare cosa essa sia veramente e come possa realmente essere di aiuto all'uomo.

Tutte le citazioni sono state tradotte dall'autore, che si scusa per la lunghezza di alcune di esse, indispensabile per sottolineare il pensiero di autori spesso citati, ma ben più raramente conosciuti.



## PREFAZIONE

L'Intelligenza Artificiale irrompe nella vita di tutti noi prendendo posizione negli smartphone e nei computer, simula i nostri comportamenti, guida le auto, risponde al telefono, si trova ormai spesso dove tradizionalmente ti aspetteresti di vedere un uomo. Siamo di fronte a processi tecnologici esplosivi che, fondati nei progressi scientifici degli ultimi decenni, sembrano guidati prevalentemente da forti interessi economici. Ma cos'è effettivamente l'Intelligenza Artificiale? È un settore dell'informatica? È una speciale tecnologia, come quella dei frigoriferi, della radio e della televisione? No, sostiene l'autore, l'Intelligenza Artificiale non si può semplicemente relegare nel dominio dell'informatica e va ben oltre la cornice della tecnologia.

I suoi lavori potrebbero essere la scintilla per innescare profondi ripensamenti, far vacillare le certezze di chi riduce il campo della filosofia alla Filosofia della Mente e al "*mind-body problem*". È infatti l'oggetto stesso dell'Intelligenza Artificiale a giustificare l'interesse per una cornice filosofica; il Pensiero è da sempre oggetto di interesse dei filosofi, che si sono già scontrati su scivolosi paradossi in alcun modo dissolti dai formalismi logici e matematici che oggi fondano l'Intelligenza Artificiale.

Per l'autore, l'intreccio è così forte che l'Intelligenza Artificiale risulta "un campo di esplorazione intellettuale che la Filosofia non può ignorare, né semplicemente studiare epistemologicamente, come si fa per altre scienze come la fisica o la biologia". Non si discute qui dunque di "atti filosofici" per analizzare il metodo scientifico utilizzato, ma si indaga sul Pensiero attraverso veri e propri studi filosofici stimolati da nuovi possenti schemi logici e matematici. C'è, insomma, uno spazio di esplorazione straordinario, che non si coglie pienamente nei laboratori di ricerca degli ingegneri.

ri e dei neuroscienziati. Questi laboratori sono tipicamente concentrati su specifici aspetti computazionali e su strutture materiali, con finalità precise. Si analizzano i segnali e l'informazione supportata, elementi che certamente incrociano il pensiero a cui, tuttavia, si scivola attorno, senza mai del tutto afferrarlo.

L'Intelligenza Artificiale, dice l'autore, "non ha oggi la filosofia che si merita". Non solo è difficile non sottoscrivere questa tesi, ma pare davvero il momento di andare oltre l'inadeguatezza di un pensiero prevalentemente sospinto da meri benefici economici, che spesso selezionano e premiano soprattutto specifici approcci scientifici funzionali a logiche di mercato. La spinta più pressante per i ricercatori viene principalmente da confronti, ormai spesso sterili, su "benchmarks" che circoscrivono in modo spesso discutibile lo studio dell'emergenza di funzioni cognitive artificiali. Si pongono obiettivi a breve scadenza, quando invece siamo di fronte a temi che richiedono verosimilmente un ripensamento profondo. Per esempio, gli studi sulla percezione, soprattutto sulla visione artificiale, raramente coinvolgono il tempo e, quando avviene, tipicamente questo non gioca il ruolo primario che si vede in altre scienze, come la fisica. Domina prevalentemente la statistica, ma si trascura la possibilità che l'interazione con l'ambiente obbedisca a vere e proprie leggi guidate dal tempo attraverso processi causali. L'approccio statistico, certamente fruttuoso anche in "computer vision", pare in realtà enfatizzare l'aggettivo "artificiale" negli associati processi intelligenti. Il riconoscimento ha prevalentemente luogo a livello di singolo frame, mentre si trascura la struttura indotta dalle regolarità temporali. Ne conseguono meccanismi cognitivi che catturano solo regolarità spaziali nelle immagini affrontando così un problema che è verosimilmente più difficile di quello offerto dalla Natura. Quando poi si ha l'ardire di emulare artificialmente il pensiero allora si valicano anche gli aspetti epistemologici necessari per formulazioni più naturali ed eleganti. La filosofia ha in serbo tradizioni secolari di studi



che intersecano innegabilmente gli attuali percorsi dell'Intelligenza Artificiale.

L'irruente invasione dell'Intelligenza Artificiale nella vita di tutti i giorni fa certamente "rientrare dalla finestra questioni che la Filosofia stessa aveva cacciato dalla porta". Nell'avvolgente divulgazione di informazioni sugli studi e sui risultati dell'Intelligenza Artificiale si inciampa spesso in tesi dove il confine tra la scienza e la fantascienza è varcato senza ritegno. Uno potrebbe farsi venire legittimamente almeno il dubbio che al computer manchi l'intenzionalità, semplicemente perché, secondo la biologia, il computer non è materia vivente. Mancano la vita e la morte, due ingredienti di non poco conto quando la sfida si concentra sull'acquisizione del pensiero. Nel palcoscenico di Shannon regna l'informazione, la simulazione dell'intelligenza che si esporta indipendentemente dal corpo. Gli attori di quel palcoscenico non muoiono, ma recitano schemi cognitivi senza timore delle ingiurie del tempo. Quegli attori hanno la forma del software, che ha raggiunto l'ambito divino obiettivo di conquistare l'infinito. È in quest'astrazione, già presente nelle idee originarie di Turing, che si nasconde l'inganno sulla presupposta conquista delle emozioni, da non confondersi né con la loro raffinata emulazione, né con il loro preciso riconoscimento. È il caro tema dell'esistenza, su cui i filosofi si sono confrontati da sempre, a posizionare l'Intelligenza Artificiale su un pianeta vicino, ma diverso dal nostro. È in questa ottica, sostiene l'autore, che "l'Intelligenza Artificiale è la continuazione della Filosofia con altri mezzi". È da quest'angolo di tradizioni di studi radicati nei secoli che meglio si collocano gli straordinari successi scientifici e tecnologici dell'Intelligenza Artificiale; l'intelligenza si presenta con le sue circolarità, con i suoi paradossi, scivolando alla presa degli scienziati come un'anguilla. L'autore affronta gli attributi dell'intelligenza; "artificiale" rispetto a "naturale", "astratta" oppure "concreta", ne promuove una collocazione filosofica secondo la quale l'intelligenza precede ogni suo attributo,

manifestando proprietà intrinseche. Questo contribuisce a turbare le strutture metodologiche stesse dell'Intelligenza Artificiale, spingendo verso interpretazioni che vanno oltre la natura biologica e il silicio dei semiconduttori. Forse è importante pensare a leggi dell'intelligenza, la cui generalità si addica a circolare agilmente tra il naturale e l'artificiale. L'autore avverte anche che lo studio dell'intelligenza potrebbe sfuggire agli schemi epistemologici abituali; la circolarità tra l'oggetto di indagine e il soggetto che esplora è un dilemma ben conosciuto in filosofia, esso affascina e inquieta allo stesso tempo. Il rilievo dell'attività neurale negli animali e negli umani, l'analisi dei segnali di un computer, conducono a uniformare i "processi mentali", ne suggeriscono la collocazione sotto lo stesso ombrello; prima o poi, si congettura, quei segnali elettrici che si vedono nei neuroni del cervello di un bambino, saranno riprodotti artificialmente. Si inciampa necessariamente nella dualità Cartesiana, nella ricerca della sede del pensiero, un aspetto da sempre al centro delle dispute filosofiche.

Sarebbe più facile, sostiene l'autore, "continuare a sperare che la tecnologia dia tutte le risposte, che esplorando la materia nell'infinitamente piccolo arriveremo finalmente alla verità, che il Vero possa essere 'dimostrato' logicamente o matematicamente". In fondo però, anche questa fiducia trasuda di fede, si ancora alla convinzione che prolungare gli straordinari successi del metodo scientifico verso lo studio del pensiero possa condurci a una sorta di risposta "finale". Non dovremo attendere troppo affinché agenti intelligenti, magari nella forma di robot per potervi associare in modo distintivo il corpo, possano vantare la transizione che conduce alla percezione di sé, alla sperimentazione delle emozioni che siamo abituati a riconoscere non solo nei nostri simili, ma anche in molti animali. Non dovremo attendere troppo per trovare sul palcoscenico due robot innamorati che parlano dei loro sentimenti e dei loro progetti futuri. Ma spartiranno davvero con noi il significato profondo del tempo, che prende forma della nascita e della

morte? Di certo, possiamo già interrogarci sul fatto se quei sentimenti li proveranno davvero! Saranno davvero i robot i protagonisti di progetti per il futuro, di desideri e di ambizioni, sperimenteranno l'amicizia e gli ideali che ci spingono a creare una società per aiutare i più deboli e i più sfortunati? L'autore si interroga dunque se non ci sia ancora uno spazio per posizioni più ricercate, più "umanistiche", che cerchino di mantenere la spiritualità del pensiero integrandola con l'approccio scientifico delle scienze cognitive. È qui che scaturisce la "consanguineità con la Filosofia" e il suo energico rilancio. Difficile pensare che i cruciali aspetti etici su cui rimbalzano molte decisioni e scelte tecnologiche oggi legate all'uso dell'Intelligenza Artificiale possano evitare l'essenza di secolari dibattiti che hanno già avuto luogo, magari solo in forma diversa. Certo è che si intravedono nuovi scenari sociali in cui gli umani si rapporteranno con agenti intelligenti operanti come loro "alter ego", così che ciroleranno anche relazioni ibride (uomo – alter ego) assieme a relazioni puramente artificiali (alter ego – alter ego). Sartre, in "Che cos'è la letteratura?" ci fa vedere i libri polverosi che, scoperti dopo molto tempo negli scaffali di una biblioteca, risuscitano l'autore, stabilendone il contatto con il lettore. Forse quegli agenti intelligenti risusciteranno il loro "padrone" dopo la morte, in modo straordinario, e al tempo stesso inquietante. Questo nuovo mondo sembra davvero richiedere la centralità dell'uomo, che deve essere coinvolto e non travolto, che deve contribuire a plasmare gradualmente il mondo nuovo. Non ci sono però sconti; serve umiltà e pazienza per capire e forgiare le straordinarie rivoluzioni dell'Intelligenza Artificiale. Serve acquisire consapevolezza di un mondo in straordinaria evoluzione, serve evitare l'asservimento dell'Intelligenza Artificiale all'unica mera logica del profitto. Questo libro apre un sentiero parallelo ai tipici percorsi tecnologici di discussione sull'Intelligenza Artificiale; spalanca gli orizzonti a indagini squisitamente filosofiche; risveglia, attraverso analisi raffinate, le grandi risorse sopite del pen-

siero filosofico; mette sul palco le menti più autorevoli con magistrali accostamenti dando spazio, attraverso citazioni, al loro pensiero. Questo libro prospetta una visione “umanistica” dell’Intelligenza Artificiale, in grado di elevare la spiritualità del pensiero attraverso i risultati delle scienze cognitive. Mi piace anche pensare che possa contribuire a diffondere l’importanza della conquista della consapevolezza del nuovo mondo prospettato dall’Intelligenza Artificiale ben oltre la barriera degli addetti ai lavori.

*Marco Gori*

Dipartimento di Ingegneria dell’Informazione  
e delle Scienze Matematiche, Università di Siena

## INTRODUZIONE

L'Intelligenza Artificiale è troppo importante, troppo interessante, per essere lasciata nelle mani degli ingegneri informatici. Il contenuto che tale disciplina porta con sé, la novità della sua forma, ne fanno un campo di esplorazione intellettuale che la Filosofia non può ignorare, né semplicemente studiare epistemologicamente come si fa per altre scienze come la fisica o la biologia. Ci sono ragioni profonde alla base di questa considerazione, che guida il presente lavoro, ragioni che tenteremo di giustificare in dettaglio nei seguenti capitoli. Poiché toccheremo campi e autori che nel nostro tempo, abituato alla specializzazione e alla dipartimentalizzazione, possono sembrare lontani anni luce, il lettore ha diritto a nostro avviso almeno a un filo conduttore per iniziare. Di qui l'esigenza di queste prime linee, che permetteranno anche di indicare cosa questo lavoro non è, né può, né vuole essere.

A titolo introduttivo, dunque, diremo che è l'oggetto stesso dell'Intelligenza Artificiale a giustificare l'interesse della Filosofia. Tale oggetto, per dichiarazione esplicita e fondativa dell'Intelligenza Artificiale, è il Pensiero; lo scopo (di nuovo in forma esplicita e fondativamente dichiarato) è la simulazione e la meccanizzazione delle attività mentali attraverso dispositivi elettronici genericamente denominati computer (a loro volta genericamente costituiti di parti *hardware* e *software*). Un secondo scopo, di natura più teorica, è la comprensione del funzionamento di queste attività mentali (qui l'Intelligenza Artificiale è esplicitamente definita come una branca della *Cognitive Science*). È chiaro che al di là della specializzazione dei vari "attori" (informatici, ingegneri, neuroscienziati ecc.) il progetto è molto coerente, con una serie di intenti coordinati e ben definiti.

Tutto ciò (l'oggetto dell'Intelligenza Artificiale, le modalità di lavoro previste, gli scopi pratici e teorici perseguiti) interpella robustamente le denominate "umanistiche" tradizioni, ma in particolare interpella la Filosofia. Fino a oggi questo appello ha preso la forma di una critica della possibilità stessa di una Intelligenza Artificiale (vedremo questo in dettaglio nel Cap. II), oppure di una generica richiesta di interdisciplinarietà tra cattedre universitarie di diversi dipartimenti, con l'intendimento che sarebbe proficuo "parlarsi" tra discipline diverse.

La Filosofia, insomma, si è innanzitutto "posizionata" rispetto all'Intelligenza Artificiale e questo posizionamento ha essenzialmente preso il carattere di una critica epistemologica più o meno condiscendente.

A loro volta, i testi teorici di Intelligenza Artificiale citano spesso e volentieri filosofi importanti per dimostrare una comunanza di interessi, per illustrare una prescienza in filosofia per problemi che adesso possono finalmente essere risolti "scientificamente", oppure per nobilitare un'attività altrimenti vissuta come troppo tecnica e specialistica. Dopo esserci occupati delle questioni della materia, sembrano dire gli scienziati, adesso con gli stessi metodi ci occuperemo anche del mentale e le premonizioni dei filosofi ci potranno essere utili per capire da dove iniziare.

Questo reciproco interesse, quando c'è stato, è rimasto appunto dunque un rapporto, un dialogo instaurato tra due elementi diversi che si parlano, comunicano, si capiscono magari, ma restano diversi per natura, scopi e ambiti di lavoro. Gli ingegneri e i neuroscienziati lavorano nei loro laboratori, fanno i loro esperimenti, proseguono le loro ricerche senza preoccuparsi di fondamenti filosofici o simili definizioni "astratte"; a loro volta i filosofi leggono e osservano quello che i primi fanno, gli epistemologi cercando di stabilirne i limiti o i reali progressi mentre i filosofi morali o metafisici cercano di dimostrare perché quanto fatto resta sempre al di sotto delle aspettative.

È tempo invece a nostro avviso per la Filosofia di riconoscere l'Intelligenza Artificiale per quello che è, cioè *se stessa*. Quello che von Clausewitz diceva di politica e guerra, sentiamo di poterlo adattare come segue:

“L'Intelligenza Artificiale non è che la continuazione della Filosofia con altri mezzi. L'Intelligenza Artificiale non è dunque solamente un atto filosofico, ma un vero strumento della Filosofia, un seguito del procedimento filosofico, una sua continuazione con altri mezzi.”

Su cosa si fonda tale convinzione, che si tratterà di sostanziare nei prossimi capitoli? Semplicemente sul fatto che la Filosofia è sempre stata una ricerca su cosa sia e come funzioni l'intelligenza umana. Qualunque interrogativo ontologico, etico, gnoseologico ha sempre dovuto fare i conti con il significato di termini come “pensare”, “conoscere” e simili. A partire dall'appercezione trascendentale kantiana il legame tra “essere” e “pensiero” è diventato addirittura esplicito, ma già in Parmenide esso appare in tutta la sua pregnanza. L'Intelligenza Artificiale in quanto ricerca sul Pensiero, oppure anche solo come tentativo di replica di esso, si propone (e filosoficamente parlando si *ripropone*) di capire proprio questo legame e le sue implicazioni, cosicché sentiamo di poter affermare che *si deve considerare l'Intelligenza Artificiale come un “essere e fare Filosofia”*.

Questo permette anche di abbozzare una definizione dell'acronimo IA senza considerazioni su una tecnologia perennemente in evoluzione. Oggi l'Intelligenza Artificiale è una etichetta che si incolla a molte attività con legami non sempre facili da distinguere. L'unità di queste ricerche non può venire dalla tecnologia, ma dallo scopo che le accomuna (pensare il Pensiero); lo stesso vale per la storia dell'Intelligenza Artificiale, che deve andare al di là dei progressi tecnici (vedremo questo aspetto in dettaglio nel Cap I). Che si tratti di sistemi esperti, di interfaccia uomo-macchina, di algo-

ritmi euristici o di *Fuzzy Logic*, tutto ciò è Intelligenza Artificiale nella misura in cui contribuisce, è guidato, conduce, a una comprensione della natura del Pensiero. E in questo si ribadisce la proposizione iniziale dell'Intelligenza Artificiale come "essere e fare Filosofia".

Questo approccio è possibile e a nostro avviso necessario. Ribadiamo che non si intende qui proporre una epistemologia dell'Intelligenza Artificiale, non fosse che per i limiti tecnici di chi scrive; né pensiamo che spetti alla Filosofia determinare se l'Intelligenza Artificiale sia o no possibile (in versione *strong* o *weak*); il passato insegna che la scienza empirica procede secondo le sue proprie vie e che la Filosofia è stata spesso smentita quando ha preteso di anticipare i tempi. Il nostro scopo è semplicemente quello di far emergere a livello filosofico e non tecnico la propria inadeguatezza della Filosofia di fronte a quello che lo spirito dell'uomo produce oggi di più avanzato.

Ma perché il rapporto possa "risolversi" adeguatamente il cammino non può essere unilaterale. È necessario andare dall'altra parte, assumere il punto di vista che sta di fronte e farlo proprio. Ora è un fatto che nel corso delle loro ricerche gli scienziati di Intelligenza Artificiale si sono presto imbattuti nei concetti/problemi di "coscienza", "intenzionalità", "realtà esteriore e interiore" e simili. Una prima reazione, fondata sull'epistemologia russelliana e favorita dalla filosofia analitica, è stata quella di semplicemente ignorare tali problematiche, tralasciandole in quanto "non definibili in termini matematici", oscure, non degne di figurare in una ricerca scientifica. Questa forma di pigrizia filosofica però non può a lungo soddisfare il ricercatore genuinamente interessato alla verità; mostreremo come proprio dalle voci più esigenti dell'Intelligenza Artificiale si levi ormai la richiesta di una risposta filosofica adeguata a quanto si fa e si scopre nei laboratori. Tenteremo di evidenziare, insomma, che *l'Intelligenza Artificiale non ha a oggi la filosofia che si merita* e che i duemila anni di ricerca filosofica che



ci precedono possono rivelarsi utili sotto vari aspetti. O, per dirla in un altro modo, siamo convinti che l'autentico ricercatore di Intelligenza Artificiale non potrà non riconoscere che determinati interrogativi tipici della storia della filosofia sono *identici* (non analoghi, non simili, ma identici) a quelli che la sua stessa pratica di ricerca gli propone.

Questo libro è dunque la testimonianza di una duplice inadeguatezza: inadeguatezza della Filosofia di fronte alle scoperte e alle promesse dell'Intelligenza Artificiale, e inadeguatezza dell'Intelligenza Artificiale di fronte alla qualità delle domande che essa stessa si pone. Se questa duplice negatività farà nascere qualcosa di positivo, starà al lettore ovviamente stabilirlo.



# 1 FILOSOFIA DELLA STORIA NELLA (E DELLA) INTELLIGENZA ARTIFICIALE

È giusto esigere che la storia, quale che ne sia l'argomento, racconti i fatti senza parzialità, senza pretendere d'avvalorare interessi o scopi particolari. Ma tale esigenza è un luogo comune, che approda a ben poco, giacché la storia di un argomento è necessariamente collegata all'idea che ci facciamo di esso. Questo fissa già in precedenza che cosa si considera importante e conveniente per l'argomento prescelto...

Hegel, *Lezioni sulla storia della filosofia, 1821-1831*

I manuali di storia dell'Intelligenza Artificiale descrivono una disciplina che nasce intorno agli anni '50, conosce un primo entusiasmo fino agli anni '80, perde il proprio slancio (*The AI Winter*) e torna in auge all'inizio del nuovo secolo; oggi applicazioni di Intelligenza Artificiale sono annunciate in qualunque settore industriale.

Che enormi progressi siano sopravvenuti è indubbio, soprattutto per quel che riguarda l'interfaccia con l'uomo (riconoscimento vocale, visione ecc.), ma a uno sguardo storico autentico questi aspetti non possono apparire sostanziali e, anzi, abbisognano essi stessi di una giustificazione. Un percorso fatto di scoperte o invenzioni casuali non è storia, è esso stesso casualità. Per capire realmente occorre una storia "filosofica" dell'Intelligenza Artificiale (se il termine sorprende parliamo di disamina concettuale); occorre considerarla per quello che essa è a livello delle idee, una gigantesca introspezione dell'intelligenza su se stessa – ossia un proseguimento dell'avventura chiamata Storia della Filosofia.

Dal punto di vista di chi crede nella filosofia, inoltre, non può che riconfortare il ritrovare temi a lui familiari in una disciplina come l'Intelligenza Artificiale che fa dell'intelligenza il suo oggetto e il suo credo. Se l'Intelligenza Artificiale è Filosofia, insomma, la sua storia deve poter essere scritta filosoficamente, ed è quello che tenderemo in questo capitolo.

*Quid* allora della tecnologia, che agli occhi di quasi tutti (inclusi i ricercatori impegnati nell'Intelligenza Artificiale stessa) è l'elemento sostanziale dell'Intelligenza Artificiale? Come negare che sia questo il lato dal quale si deve *in primis* considerare la storia dell'Intelligenza Artificiale? La questione merita un approfondimento.

La storia canonica ci dice che l'Intelligenza Artificiale inizia con la geniale intuizione di Turing, che vede nel computer (un successo ingegneristico dell'epoca) la promessa di una potenza elaborativa senza precedenti. Si apre la strada alla computazione universale, la *Universal Turing Machine* come macchina interamente programmabile è possibile. I segni e sogni precedenti finalmente si avverano, dopo la teoria viene la pratica. Ed è la potenza del digitale che permette tutto questo, dunque un miglioramento tecnico. L'elettronica è un aspetto secondario per Turing che ammette tranquillamente la possibilità di altre tecnologie, ma comunque secondo la storia ufficiale e quasi tutti i manuali sull'argomento all'inizio c'è un artefatto, una macchina, il computer elettronico digitale.

“Siamo tanto più pronti a farlo se si considera il fatto che l'odierno interesse in macchine che pensano è stato suscitato da un particolare tipo di macchina, abitualmente chiamato “computer elettronico” o “computer digitale”. Seguendo questo suggerimento permetteremo solo a computer digitali di prendere parte al nostro gioco”<sup>1</sup>.

<sup>1</sup> A. M. Turing, *Computer Machinery and Intelligence*, Mind 49, 1950, Oxford UK, Oxford University Press.

Questo inizio è però arbitrario né più né meno di altri. Tecnicamente si può far partire il racconto con Charles Babbage e il suo *Analytical Engine*; teoricamente con Leibniz e la *Characteristica universalis*; metodologicamente con Frege e la *Begriffsschrift*.

Ciò che si coagula in Turing non è l'artefatto – il computer digitale – ma la domanda: “*Can machine think?*”

È la formulazione di questa domanda che fa di Turing il padre fondatore dell'Intelligenza Artificiale. L'interrogativo appare provocatoriamente nella prima riga di “*Computer Machinery and Intelligence*”, viene ritirato esattamente un paragrafo più tardi per poi risorgere a metà del discorso e continua a galleggiare nel retroterra dell'intera argomentazione senza mai essere esplicitamente accettato né rifiutato<sup>2</sup>.

Questa domanda è poi definitivamente adottata dai partecipanti alla conferenza di Dartmouth nel 1956, dove viene coniato il termine “*Artificial Intelligence*”. Siamo in un'epoca dominata dal logicismo, dalla filosofia analitica, dal neopositivismo; la metafisica è considerata un fumoso chiacchiericcio, le questioni una volta considerate fondamentali sono ridotte nel migliore dei casi a problemi di linguaggio. È probabilmente la sola epoca nella quale una domanda del genere poteva essere formulata in questi termini. Ma con questa domanda – e non con l'arrivo del dispositivo “computer” – l'Intelligenza Artificiale si eleva oltre l'orizzonte della tecnologia, va oltre le proprie intenzioni: nella tecnicità stessa della domanda, nell'atto stesso in cui il Pensie-

<sup>2</sup> Ci sarebbe molto da analizzare a livello stilistico su quel testo fondatore dell'Intelligenza Artificiale che è *Computing Machinery and intelligence* e su quelli che lo precedono. Turing, matematico brillante e provetto ingegnere, si dimostra qui anche abile propagandista, includendo nella discussione varie obiezioni alla sua tesi e dichiarando allo stesso tempo che la tesi non ha senso (*meaningless*) anche se potrebbe averne in futuro. Tutto questo ha come effetto che ancora oggi esistono varie interpretazioni su che cosa Turing esattamente sostenesse nei suoi scritti; per maggiori dettagli rimandiamo alla lettura dell'eccellente contributo di Shlomo Danziger, in “Philosophy and the theory of AI 2017”, Springer 2017.

ro è fatto oggetto di studio alla stregua della Natura, si risuscitano proprio quelle secolari questioni ormai dichiarate indifferenti. Fissare l'inizio dell'Intelligenza Artificiale all'intuizione di Turing è adeguato, corretto; proprio questo qualifica l'Intelligenza Artificiale come un'avventura filosofica, dove il ruolo della tecnologia è preponderante ma non fondante. L'Intelligenza Artificiale, insomma, non è una tecnologia, non lo è mai stata, fin dai suoi inizi.

Se si crede nella Filosofia, infine, bisogna essere grati all'Intelligenza Artificiale; grazie a essa rientrano dalla finestra questioni che la Filosofia stessa aveva cacciato dalla porta (vedi *Parte I, Elementi di filosofia nella storia dell'Intelligenza Artificiale*). Si può anche dire che con l'Intelligenza Artificiale la Filosofia ritrova un contatto autentico con se stessa dopo un prolungato periodo di stasi intellettuale (vedi *Parte II, In cammino verso l'Intelligenza Artificiale*). Ma si sta in fondo dicendo la stessa cosa.

## ELEMENTI DI FILOSOFIA NELLA STORIA DELL'INTELLIGENZA ARTIFICIALE

### *Osservazione, empirismo, materialismo*

La nostra idea e concezione della scienza, di qualunque scienza in quanto tale, si accompagna inevitabilmente allo strumento, supporto indispensabile dello scienziato e componente essenziale di qualunque laboratorio di ricerca.

Da Galileo in poi la scienza empirica ha sempre avuto un rapporto privilegiato con lo strumento scientifico. L'innovazione principale dello stesso Galileo non è di natura nozionale (molte delle sue conclusioni si sono rilevate errate) ma consiste nella teorizzazione dello strumento come legittimo mezzo di ricerca. Se gli

oggetti nello spazio sono difficili da osservare a occhio nudo, allora si deve costruire uno strumento per “avvicinarli” alla nostra osservazione; i nostri sensi possono e devono essere potenziati così da permetterci di guardare al di là delle loro limitazioni.

Nelle diatribe del grande pisano con la Chiesa sono documentate obiezioni di vari teologi che si rifiutavano di guardare attraverso il telescopio perché non si fidavano di quello che lo strumento rivelava, non esistendo nessuna garanzia del fatto che l'osservazione mediata da uno strumento fosse conforme alla realtà. Ma queste obiezioni (non sprovviste di una loro coerenza come vedremo in seguito) non hanno frenato la produzione e l'utilizzo massiccio di strumenti sempre più elaborati.

Microscopi, riflettori, acceleratori nucleari obbediscono tutti alla stessa logica: i nostri sensi sono limitati e ci servono mezzi per osservare la materia (o la realtà) in modo migliore e più dettagliato. Anche la neurofisiologia obbedisce a questa logica, disegnando “mappe” del cervello e localizzando le varie funzioni della memoria, del linguaggio ecc.

L'Intelligenza Artificiale invece, sempre nell'atto fondativo di Turing, inizia come noto con la simulazione, stilizzata nell'“*imitation game*”: se un dispositivo riesce a impersonare un essere umano in maniera convincente possiamo definirlo “intelligente”. C'è una forte componente anti-essenzialistica in questo approccio, che dichiara accontentarsi del fenomeno senza interessarsi alla “cosa-in-sé”.

Il metodo “simulativo” cozza contro tutta la tradizione scientifica precedente: il mentale non è osservabile a occhio nudo, come non lo erano gli atomi all'epoca di Galileo. Ma mentre Galileo per prima cosa costruisce “un occhialino per vedere le cose minime”<sup>3</sup> al mentale questa possibilità non viene offerta. Questo

<sup>3</sup> G. Galilei, *Lettera a Federico Cesi*, Firenze, 23 settembre 1624, in G. Galilei, *Opere*, ed. nazionale a cura di A. Favaro, Firenze 1968, vol. XIII, pp. 208-209.

anti-essenzialismo è pesantemente presente nella teoria dell'Intelligenza Artificiale, rassicura i ricercatori e li convince di fare scienza, ingegneria, qualcosa di concreto, non filosofia<sup>4</sup>. A una più attenta riflessione si tratta però di una flagrante ammissione della natura anche “metafisica”<sup>5</sup> dell'Intelligenza Artificiale. Chi si accontenta del fenomeno riconosce che dietro esso impera e necessariamente resiste un noumeno, una essenzialità, magari irraggiungibile ma presente.

La filosofia ha vissuto questa tragica situazione almeno da Kant in poi nei lunghi dibattiti sulla cosa-in-sé. Questa natura metafisica può e deve essere accettata per quello che è, non esorcizzata con continue – e infondate – professioni di fede all'empirismo o alla materia quale realtà ultima dell'universo. Convivere con la metafisica, lottare con essa, sopravviverele, la filosofia lo fa da sempre: ecco una lezione che una Intelligenza Artificiale consapevole di se stessa deve poter assimilare.

### *Software, significato, linguaggio*

Al momento della conferenza al Dartmouth College (1956) per “computer” si intendeva un dispositivo programmabile, cioè capace di ricevere determinate istruzioni in un determinato linguaggio appositamente creato e di eseguire altre istruzioni secondo regole prestabilite. Queste istruzioni potevano essere correlate da informazioni raccolte dall'ambiente oppure inserite dal programmatore (*database*): tutto comunque deve essere scritto nello stesso linguaggio. “Linguaggio appositamente creato” significa ovviamente linguaggio di programmazione e costituisce

<sup>4</sup> Le pagine che John McCarthy dedica a queste riflessioni (che vedremo in dettaglio nel prossimo capitolo) sono rivelatrici al proposito.

<sup>5</sup> Qui il termine “metafisica” può essere compreso nel suo significato tradizionale (ricerca che va oltre la natura fisica) o nell'accezione popperiana di non falsificabilità. L'affermazione vale a nostro avviso in entrambi i casi.



l'intero universo del computer, poiché qualunque informazione gli giunga dall'esterno dovrà sempre essere tradotta per poter essere compresa (nel senso di com-prendere, cioè prendere insieme al resto).

Questo particolare rende determinante l'aspetto della programmazione (il software) rispetto all'hardware, poiché il senso (il significato) di qualunque istruzione o informazione ricevuta o inviata è determinato da esso. Per questa ragione molti teorici dell'Intelligenza Artificiale tendono a concentrare la loro attenzione sui programmi piuttosto che sull'hardware; e, seguendo il modello simulativo di cui sopra, sempre per questa ragione c'è stata una tendenza diffusa a considerare il software come l'equivalente della mente e l'hardware come l'equivalente del cervello. L'argomento è il seguente: il software, al di là della estraneità del linguaggio per un essere umano, contiene un significato, che il computer è capace di comprendere. L'essere umano non lo capisce (poiché è come se fosse scritto in una lingua straniera), ma la macchina sì; dalla comprensione viene l'azione di risposta, adeguata o meno. Il problema della materialità del programma (elettricità e *byte*) è indifferente, anche il linguaggio umano ha una sua materialità (suoni, scrittura) che non inficiano il fatto che al di là del fisico ci sia un senso. Il software è la semantica del computer; che tale semantica ci sia è dimostrato dalla reazione stessa del computer (*l'output*), questo è ciò che permette di affermare "la macchina pensa", anche se è un pensare diverso da quello umano. Tutto questo dovrebbe bastare a convincere qualunque persona ragionevole che è possibile immaginare una macchina che pensa.

Prendiamo un caso diverso, un uomo al comando di una ruspa azionata da leve per alzare, abbassare, ruotare la scavatrice. L'uomo abbassa la prima leva, e la ruspa si abbassa, poi la alza, e la ruspa si alza. In cosa si differenzia questa operazione da un comando dato a un computer? Il gesto "alzare la leva" è l'equivalente di

una istruzione, cui la ruspa risponde alzando la scavatrice; certo in questo caso la leva ha un effetto materiale diretto e fisico sulle diverse parti connesse del dispositivo fino all'estremità che si solleva, ma la differenza non può essere nella materialità del gesto poiché anche nel computer tale materialità esiste, in forma di impulsi elettrici.

Per quel che riguarda la differenza di linguaggio, è evidente che anche qui l'uomo compie una traduzione: l'intenzione di alzare la scavatrice viene tradotta nell'alzare la leva (che non è la stessa cosa) per permettere al dispositivo di recepire l'istruzione in un linguaggio a esso comprensibile. Eppure nessuno sosterebbe che la ruspa "pensa", anche se esegue l'istruzione alla perfezione. Ovviamente il computer è programmabile, mentre per quante funzioni si aggiungano alla ruspa queste resteranno sempre di natura puramente fisica.

Fermiamoci un momento: la programmabilità del computer significa allora che entrano in gioco elementi non più solamente "fisici"? Che rapporto c'è allora tra la natura dell'algoritmo (del software) e la materia? E come interagiscono tra loro? Attenzione, già stiamo parlando di filosofia!

Andiamo oltre: immaginiamo un programma informatico che contenga le varie istruzioni che governano la ruspa (se la leva è abbassata, allora spostare il braccio meccanico verso il basso ecc.); inseriamo un computer così programmato tra noi e la ruspa. Poi diamo al computer il controllo della ruspa e azioniamo il comando "leva alzata". Accadrà lo stesso che nel caso precedente, avremo ora una ruspa "digitale", ma se diciamo che il computer pensa in realtà stiamo dicendo che noi dall'esterno abbiamo creato una realtà simbolica (un doppione del mondo raccontato con un linguaggio diverso, un software) che duplica la meccanicità delle azioni della ruspa in un universo parallelo. Questo magari facilita il lavoro dell'operaio alla ruspa, ma l'interrogativo sulla natura del secondo universo resta; senza una spiegazione su questo dovremo ammet-

tere che non c'è traccia di “comprensione”, altrimenti dovremmo sostenere che anche una palla da biliardo “capisce” che deve spostarsi quando è colpita dalla stecca.

Questa discussione (riassumibile a livello generale nell'affermazione che software e hardware sono come mente e cervello) ci porta inevitabilmente alla cartesiana distinzione *res cogitans* / *res extensa*; che il problema di tale distinzione abbia una diretta influenza anche su cosa significa “significato” mostra che non basta dichiarare la metafisica e l'ontologia dei non-problemi per risolvere la questione.

### *Pensiero, sensazioni, intelletto*

Il modello simulativo (con il computer) ignora invece completamente le intuizioni o sensazioni. Esse, come il pensiero, non sono osservabili, eppure l'Intelligenza Artificiale discute di macchine intelligenti, che pensano, mai di macchine senzienti. C'è da chiedersi perché anche i più ottimisti per quel che riguarda la possibile duplicazione del cervello non abbiamo mai sentito lo stimolo per un approccio analogo al sistema nervoso, che non sembra a priori fisicamente più complesso<sup>6</sup>.

La ragione sembra essere che il computer è stato immediatamente e quasi naturalmente situato al livello dell'Intelletto (*Verstand*), nel senso funzionale definito da Kant. L'Intelletto kantiano è quella facoltà che determina, definisce, fissa la conoscenza in ambiti specifici e separati, insomma è la funzione che gestisce i concetti. Ma già Kant avvertiva che “i concetti senza le intuizioni sono vuoti”, molto si è tentato per riempire quel vuoto.

<sup>6</sup> Si potrebbe obiettare che esiste un'avanzata ricerca sulle reti neurali, che indubbiamente fa parte dell'Intelligenza Artificiale. Ma tale ricerca è quasi esclusivamente indirizzata al *Machine Learning* – che discutiamo più avanti – non alla duplicazione del sistema nervoso. Rimandiamo al Cap. IV per un'analisi di dettaglio di come l'Intelligenza Artificiale più recente abbia affrontato la questione delle attività senzienti dell'essere umano.

Oggi diciamo naturalmente che i computer sono in grado di “vedere” o “ascoltare”; in realtà questo significa semplicemente che un programma è in grado di “tradurre” in linguaggio simbolico alcuni dati della realtà esterna cui ha accesso. In altre parole, si tratta di nuovo della creazione di un universo parallelo gestito da un linguaggio appositamente creato. Nella mentalità generale dell’Intelligenza Artificiale le sensazioni esistono non tanto in quanto pensiero, ma in quanto sono formalizzabili. E anche qui tocchiamo un tema (i rapporti tra percezione, intuizione, concetti) che non si può non definire filosofico.

Lo stesso dicasi per le modalità di ragionamento del computer: alla logica di primo livello si è presto affiancata la *Fuzzy Logic*, i sistemi modello *Pandemonium*, le reti bayesiane nelle loro varie forme, la logica non-monotonica. Dopo la macchina fredda e calcolatrice, insomma, si è cercato di introdurre le incertezze della vita reale, seguendo il filo conduttore della maggiore flessibilità e del calcolo probabilistico. Il mero “dato” deve adesso essere ponderato, giudicato, pesato in funzione di altri elementi più vicini agli interessi e alle inclinazioni dell’uomo. Dopo la rigidità del giudizio determinante, si cerca di riprodurre ora la facoltà del giudizio riflettente, dove il “dato” intuito non ha un universale già esistente (una categoria) e anzi tale universale è da ricercarsi<sup>7</sup>.

Dopo il pensare puro, ci si occupa del fare. Il singolo computer, la macchina *stand-alone* cede il passo alle architetture di rete, dove quello che conta è la connessione; e l’intelligenza si va ora a cercare (se non addirittura a definire) nella collaborazione tra macchine connesse, poi tra macchine ed esseri umani e, infine, tra oggetti

<sup>7</sup> C. F. Kant, *Critica del Giudizio*, 1790. Questa ricerca delle intuizioni come necessario pendant dei concetti sarà come vedremo alla base di una larga parte dei lavori di John McCarthy – il creatore del termine *Artificial Intelligence* – tra altri. Per ora ci interessa solo sottolineare quanto il kantismo sia presente nelle impostazioni dell’Intelligenza Artificiale e come le ricerche nel bene e nel male ne risentano.

di uso umano (frigoriferi, scarpe, automobili) che vengono definiti “intelligenti” in quanto inseriti nella logica della rete che tutto collega. L'Intelligenza Artificiale si “sociologizza”, *dopo la ragion pura c'è la ragion pura pratica*, l'Intelletto kantiano esce dalla sua autoreferenzialità ed entra nel mondo.

Non tutta l'Intelligenza Artificiale ha seguito questi schemi (vedremo più tardi i lavori per esempio di Rodney Alan Brooks e la sua proposta di una “Intelligenza Artificiale robotica”), ma una gran parte sì. In filosofia non si può non fare i conti con Kant, vedremo che il post-kantismo ha per esempio ampiamente argomentato che questa distinzione tra Intelletto, Ragione e Intuizione, essendo una distinzione dell'Intelletto stesso, sia invalida o perlomeno unilaterale, comunque non adatta a coprire l'intero ambito dell'umano agire.

### *Una macchina infinita*

Per concludere la nostra analisi del computer, notiamo che si tratta di un dispositivo essenzialmente infinito. Indipendentemente dalla sua potenza di calcolo, sarà sempre possibile costruirne uno più potente, indipendentemente dalla sua memoria sarà sempre possibile aggiungerne dall'altra. Questo aspetto è vissuto nell'Intelligenza Artificiale come essenzialmente positivo, sia che si consideri la mente umana come un calcolatore molto potente che il computer deve solo raggiungere in termini quantitativi (la forza bruta) sia che l'aumento di memoria sia in sé promessa di poter un giorno racchiudere l'intera conoscenza in una banca dati. La potenza del digitale intravista da Turing sembra insomma portare con sé la possibilità di formalizzare l'intero universo. Ma l'infinità è un concetto scivoloso: è proprio Leibniz, che per primo ipotizza la possibilità di formalizzare l'universo, a dover concepire una infinità di mondi; e nella forma dei *counterfactuals* di David Kellog Lewis, i mondi infiniti appaiono nei classici problemi di “Intelligenza Artificiale Logica”.

## IN CAMMINO VERSO L'INTELLIGENZA ARTIFICIALE

Domandandoci infatti: che cos'è la filosofia, noi parliamo sulla filosofia. Ponendo la domanda in questi termini, ci collochiamo in una zona che si trova al di sopra e quindi al di fuori della filosofia. Ma lo scopo della nostra domanda è piuttosto quello di penetrare nella filosofia, di prendervi dimora e di comportarci nel modo che le è proprio, vale a dire di filosofare. Il cammino del nostro colloquio non deve perciò avere soltanto una direzione chiara, ma deve al tempo stesso far sì che tale direzione ci dia la certezza di muoverci all'interno della filosofia e non di girarvi intorno restandone fuori.

Martin Heidegger, *Che cosa è la filosofia*, 1956

### *Computer: il mezzo e il fine*

Abbiamo già visto come il computer elettronico digitale non possa essere assimilato a uno strumento scientifico tradizionale. Esso non serve a visualizzare meglio la realtà al di là dei sensi, è il suo comportamento nel mondo a contare secondo il modello simulativo. In questo senso la macchina è il *mezzo* per provare una determinata teoria (diversamente dalle scienze tradizionali dove la prevedibilità o comportamento da verificare è delle cose, non dello strumento). Ma la macchina è anche lo *scopo* dell'Intelligenza Artificiale, poiché si tratta di un artefatto che deve essere costruito dal ricercatore. A causa del modello simulativo, la macchina è il mezzo della ricerca e lo scopo della ricerca, idealmente entrambe le cose. Ecco perché in Intelligenza Artificiale si parla di test e non di prova o esperimento: un test si può fare *al* computer, *sul* computer o *con il* computer, il che lascia le cose abbastanza nel vago da non doversi soffermare su questo punto. In realtà ogni test testa se stesso, non una teoria generale.

In chiave di questo doppio status si può per esempio riformulare l'intero dibattito sul ruolo della rappresentazione in Intelligenza Artificiale (il problema della *Knowledge Representation*). È uno dei temi più astratti, ed è solitamente affrontato in chiave epistemologica: come deve rappresentarsi il computer il mondo esterno? Meglio una rete neuronale o un sistema simbolico? E come fondare il significato dei simboli sulla realtà fisica che deve essere rappresentata? Oppure, ha senso l'idea stessa di rappresentazione? Nel prossimo capitolo analizzeremo il versante epistemologico della discussione, qui ci interessa solo riformulare il problema.

Se quello che interessa è lo scopo (in termini filosoficamente precisi, se lo scopo è la verità del mezzo) la rappresentazione è considerata fondamentale. Prendiamo il caso di per esempio John Mc Carthy e il suo *designer stance*.

“Tuttavia, su questi concetti l'Intelligenza Artificiale adotta una posizione che potremmo chiamare da designer; si chiede di che tipo di conoscenza, credenze ecc. ha bisogno un computer per comportarsi in modo intelligente e come inserire tali conoscenze in un programma informatico. La posizione del designer guarda un artefatto o un organismo esistente in termini di ciò che è progettato per fare o si è evoluto per fare. Essa studia come progettare un artefatto. Ciò può richiedere il conferire a esso conoscenza, credenze ecc. e la capacità di pianificare ed eseguire piani”<sup>8</sup>.

L'ultima frase è rivelatrice: spiega l'impegno quasi ossessivo di Mc Carthy fin dal 1959 per una formalizzazione del *common sense*, della realtà quotidiana<sup>9</sup>.

<sup>8</sup> John McCarty, *Philosophy of AI and AI of philosophy*, p. 3, 2006 (corsivo dell'autore).

<sup>9</sup> “La posizione metodologica di McCarthy non è sostanzialmente cambiata rispetto alla sua prima articolazione in ‘McCarthy 1959’, solo elaborata e arricchita in ‘McCarthy & Hayes 1969’. La motivazione per usare la logica è che – anche se l'eventuale implementazione non utilizza direttamente e semplicemente tecniche di ragionamento logico come la dimostrazione di un teorema – la formalizzazione logica ci aiuta a compren-