

Contributi/5

La mente esteriorizzata

Per una critica genetica dell'Intelligenza Artificiale

Lorenzo De Stefano  0000-0003-3339-2541

Articolo sottoposto a *double-blind peer review*. Inviato il 16/03/2023. Accettato il 20/09/2023.

THE EXTERIORIZED MIND. FOR A GENETIC CRITIQUE OF ARTIFICIAL INTELLIGENCE

The present paper aims to explore the genesis of Artificial Intelligence (AI) through the lens of Ernst Kapp's theory of organ projection. The introductory section provides an overview of the current issues surrounding AI within the contemporary debate. Subsequently, the second paragraph delves into Kapp's theory of organ projection as a comprehensive theoretical framework for the analysis. The third part of the paper entails a genealogical and phenomenological analysis of the earliest archetypical models of AI, namely the Turing Machine and von Neumann's architecture. Within this context, I examine Turing's and von Neumann's theories on computers and the brain, along with their critiques, while also considering the possibility of interpreting these machines as products of organ projection. In the final concluding section, I scrutinize the strengths and weaknesses of the theory of projection as a hermeneutical and genealogical model for comprehending AI.

1. Intelligenza ed esteriorizzazione

La possibilità di creare artificialmente un'intelligenza in grado di replicare e superare le funzioni e capacità del cervello organico, ha animato sin dal secolo scorso sia il dibattito scientifico sia le narrazioni letterarie¹ e risulta tutt'oggi, dopo alterne vicende, una prospettiva di ricerca aperta e lautamente finanziata². Tecnologie come *il machine learning, deep learning*, il NLP (*natural language processing*) le ANN (*artificial neural network*), la *A-Life*, l'avvento del

¹ In letteratura celebre a tal proposito è il racconto *La risposta* di Fredric Brown, pubblicato nel 1954, che tratta della realizzazione di una macchina planetaria intelligente in grado di connettere tutto il sapere dell'universo; tale macchina realizzerebbe la creazione di Dio. Cfr. F. Brown, *Tutti i racconti (1950-1972)*, tr. it. Aa.Vv., Milano 1992.

² Per una storia dettagliata della AI dagli albori agli ultimi sviluppi si rimanda a M. A. Boden, *L'Intelligenza Artificiale*, a cura di D. Marconi, Bologna 2019.

web semantico, sono solo alcune delle ultime frontiere del AI, che sembrano avvicinare la possibilità della AGI (*artificial general intelligence*)³, vale a dire di un'intelligenza macchinica in grado di comprendere, apprendere, adattarsi e svolgere diverse attività cognitive in modo simile a un essere umano.

Questo traguardo non è ancora stato raggiunto, in quanto il modello di intelligenza artificiale oggi diffuso è ancora essenzialmente di tipo debole (*weak AI* o *narrow AI*), ossia basato su tecniche di *machine learning* e di *deep learning*, in grado di superare le prestazioni umane, ma solo in specifiche funzioni. Nonostante le tecnologie attuali non consentano ad oggi di replicare in toto la duttilità di una qualsiasi intelligenza organica⁴, la possibilità di riprodurre artificialmente l'intelligenza umana continua ad eccitare l'immaginario collettivo ben oltre la semplice fantascienza⁵.

Quotidianamente uno stuolo di critici ed apologeti dibatte se tale singolarità sia davvero a venire e realizzabile, se essa sia una idea di ragione cui la ricerca scientifica e tecnologica deve tendere, oppure se essa sia un generale *misunderstanding* basato su una fallacia teorico-epistemologica sfociante in una eguaglianza troppo semplicistica tra cervello e macchina. È quindi sul terreno epistemologico ed ontologico che la questione della AI e delle sue forme generali come la Superintelligenza⁶ e la Singolarità⁷ deve essere posto ed analizzato. A tale scopo, è necessario preliminarmente analizzare genealogicamente la nascita della AI a partire dai disegni e le convinzioni dei suoi primi teorici, dai suoi modelli

³ Acronimo di *Artificial General Intelligence*. L'AGI è un tipo di intelligenza artificiale in grado di comprendere, imparare e applicare conoscenze in diverse attività, simili a quelle umane. A differenza dell'AI specializzata, che ha compiti limitati, l'AGI mira a emulare l'intelligenza generale umana, affrontando una varietà di problemi con la stessa flessibilità e adattabilità di una mente umana. L'obiettivo principale della creazione dell'AGI è superare le capacità cognitive umane, adattandosi a nuove situazioni come farebbe un essere umano. Un'AGI dovrebbe, in teoria, apprendere da esperienze, risolvere problemi complessi, comprendere il linguaggio naturale, ragionare, esprimere creatività e possedere altre abilità intellettuali. Attualmente, l'AGI pienamente funzionale è ancora in fase di ricerca e sviluppo. Gli scienziati e gli ingegneri stanno costantemente lavorando per sviluppare tecnologie avanzate che ci avvicinino alla realizzazione di questa forma di intelligenza artificiale.

⁴ Per una aggiornata ricognizione delle applicazioni in cui l'intelligenza umana si rivela ancora superiore rispetto a quella artificiale e per una critica della razionalità algoritmica si rimanda al recente lavoro di G. Gigerenzer, *Perché l'intelligenza umana batte ancora gli algoritmi*, tr. it. R. Mazzeo, Milano 2023.

⁵ Recentemente un ingegnere di Google Blacke Lemonie ha affermato che l'intelligenza artificiale che stava istruendo, chiamata LaMDA, ha preso coscienza di sé. Le reti neurali artificiali che Lemonie stava perfezionando avrebbero dato vita a una sorta di coscienza artificiale (cfr. <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine>).

⁶ Termine introdotto da Nick Bostrom nel suo *Superintelligenza. Tendenze, pericoli, strategie*, tr. it. S. Frediani, Torino 2018) che indica un'intelligenza artificiale talmente superiore in performatività alla razza umana da metterne in pericolo l'esistenza.

⁷ La nozione di singolarità (*singularity*) è stata introdotta da Ray Kurzweil nel suo saggio *Singularity is Near* (Id., *La singolarità è vicina*, Milano 2008). Essa sarebbe il risultato dello sviluppo di genetica, nanotecnologia e robotica che rendono possibile l'ibridazione e il potenziamento dell'umano con macchine intelligenti. Dall'unione tra intelligenza umana e artificiale nasce una nuova forma ibrida che Kurzweil denomina appunto singolarità che estenderebbe il suo dominio sull'universo.

di funzionamento e dalle forme in cui essa si è sostanziata nel corso della sua relativamente breve storia. In questo articolo ci proponiamo di fornire un primo passo in questa direzione.

Sin da Aristotele, che la identificava come la facoltà di astrarre l'universale dal particolare, l'intelligenza, spesso troppo semplicemente identificata con la mera capacità razionale, ha costituito il *principium individuationis* dello *zoon logon echon*. La questione anima la filosofia sin dall'oscura riflessione di Eraclito, in cui il *Logos*, la cui polisemia rinvia al discorso, al linguaggio ma anche al tenere assieme⁸, è proprio l'insondabile intelligenza e ragione che agli uomini è sempre celata e secondo cui avvengono tutte le cose. *Logos* era la legge della *physis* con cui lo *aner philosophos* deve convenire (*homologhein*) traendo da essa la misura del suo stesso intelletto.

Nei duemilacinquecento anni che ci separano dalla filosofia eraclitea il *Logos*, che ha smesso di indicare la prerogativa cosmologica della *physis* per divenire aristotelicamente la facoltà dell'anima propria dell'animale razionale, tuttavia mantiene la sua enigmaticità; pensiero, ragione, intelligenza, ingegno (*mechanê*) sono ancora oggi termini ambigui al punto che, oltre a sottrarsi a definizioni univoche, non sono nemmeno più ritenuti fenomeni unicamente umani, ma connaturati al vivente in generale⁹, e, come vedremo, anche a un particolare tipo di enti tecnici e informazionali. Dalla Grecità ad oggi modelli e concezioni si sono succeduti lungo la storia del pensiero metafisico e scientifico occidentale. È con il razionalismo di Descartes che il *cogito* – declinazione moderna del *logos* aristotelico – diviene il principio di individuazione ontologico della soggettività. Kant poi articolerà ulteriormente l'intuizione razionalista da un lato facendo dello *Io penso* il principio supremo di tutta la conoscenza umana nella sua capacità di unificare il molteplice sensibile e di sussumerlo sotto le categorie, dall'altro, nel paralogismo trascendentale della *Critica della Ragion pura*, svincolando definitivamente l'anima da ogni funzione logica e decretandone l'indimostrabilità per la ragione teoretica. Kant scinde una volta per tutte le facoltà di ragione ed intelletto, la prima *organon* delle idee, la seconda dei concetti. Sarà poi Nietzsche ad operare il passaggio successivo verso una naturalizzazione di intelletto, coscienza e intelligenza come manifestazione della biologica e fisiologica volontà di potenza, aprendo di fatto la strada in filosofia a un approccio evuzionistico e biologistico. L'insondabilità e le difficoltà di definizione di mente, ragione e intelletto rimangono ancora oggi questioni aperte in sede di psicologia e neuroscienze, così come permane largamente dibattuto cosa sia intelligenza. Una definizione particolarmente accurata e omnicomprensiva è stata tentata da una dichiarazione editoriale di cinquantadue ricercatori nel 1994:

⁸ Per una ricostruzione dell'etimologia e del senso della parola *logos* si rimanda al saggio di Heidegger *Logos* in Id., *Saggi e Discorsi*, a cura di G. Vattimo, Milano 1976, pp. 141 ss.

⁹ Per una ricostruzione della ontogenesi e della filogenesi del comportamento intelligente si rimanda al volume di P. Godfrey-Smith, *Metazoa. Gli animali e la nascita della mente*, tr. it. I. C. Blum, Milano 2012.

A very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings – ‘catching on’, ‘making sense’ of things, or ‘figuring out’ what to do¹⁰.

Possiamo osservare come, allo stato attuale del progresso tecnologico, questa descrizione possa calzare per alcune forme di AI, in quanto sanno perfettamente ottemperare ad alcune funzioni intelligenti anche contemporaneamente come il prendere decisioni orientate sulla base di dati, mentre ad ora non esistono tecnologie in grado di imitare la capacità di astrazione e di ‘comprensione’ semantica e concettuale. Possiamo dire schematizzando al massimo che se le macchine possono ad oggi esibire comportamenti intelligenti, esse non producono conoscenza o scienza, le quali rimarrebbero prerogative essenzialmente umane¹¹.

Dunque, il terreno teorico per comprendere effettivamente che cosa consideriamo sia l’intelligenza e, in seconda battuta, cosa intendiamo quando parliamo della sua trasposizione artificiale appare non del tutto dissodato, vista la stratificazione teorica filosofica e scientifica sul tema, senza contare il dibattito biologico che rintraccerebbe la comparsa della mente sin dall’apparizione dei metazoa¹².

Ciò ci porta in via preliminare a una serie di considerazioni:

1) qualunque tentativo scientificamente ponderato che miri a inquadrare ontologicamente, epistemologicamente ed eticamente l’intelligenza artificiale deve assumere a priori una posizione su che cosa sia o cosa si intenda per intelligenza e artificio. Deve ossia aver accesso a una dimensione ermeneutica e genealogica.

2) si può parlare di intelligenza artificiale solo se si accetta una necessaria dose di riduzionismo, vista la vastità del fenomeno ‘intelligenza’ e la complessità di ciò che si può definire ‘artificiale’ in antitesi al termine ‘naturale’.

3) Solo una volta che si è risposto a tali questioni è possibile decidere se una intelligenza, o singolarità artificiale possa essere dotata di una *agency* o di intenzionalità, sia quindi sovrapponibile a quella umana.

4) Un’analisi della intelligenza artificiale deve basarsi su una filosofia e una antropologia della tecnica che ci aiutino a determinare cosa sia artificio e cosa

¹⁰ L. S. Gottfredson, *Mainstream Science on Intelligence: An editorial with 52 signatories, history and bibliography*, «Intelligence», 24 (1), Delaware 1997, pp. 13-23.

¹¹ Sul tema vedi M. Coeckelbergh, *AI Ethics*, Princeton 2020 e L. Floridi, *Etica dell’intelligenza artificiale. Sviluppi, opportunità, sfide*, Milano 2022.

¹² Cfr. P. Godfrey-Smith, *Metazoa*.

organismo, o, in generale, che rapporto sussiste tra uomo e tecnicità nel dialogo con le scienze positive.

5) La possibilità di definire una AI ‘forte’ dipende dall’apriori filosofico, esplicito o implicito, di partenza.

Il presente contributo, che è un primo passo verso un indirizzo di ricerca più vasto di lettura della AI alla luce di una teoria generale della tecnicità, tenterà di fornire risposte ad alcune questioni sopra elencate, mentre altre saranno materia per un futuro lavoro; in questa sede cercherò di analizzare la genesi della AI nella sua prima formulazione compiuta presso Turing e von Neumann dal punto di vista antropologico e genealogico, non prenderò in considerazione, pertanto, gli ulteriori sviluppi della ricerca e le ultime tecnologie. Questo perché gli approcci dei due matematici contengono alcuni degli assunti teorici fondamentali che hanno indirizzato la ricerca nel campo della AI: uno tra tutti il parallelismo tra computer e cervello. Proverò, dunque, a ricondurre in prima istanza la questione della AI all’interno di una teoria generale della tecnogenesi secondo il concetto di esteriorizzazione di Ernst Kapp; in secondo luogo, prenderò in considerazione i lavori dei padri della intelligenza artificiale, Alan Turing e John von Neumann, per evidenziare come se da un lato le loro architetture possano essere interpretate secondo il paradigma della proiezione, dall’altro l’equiparazione tra computer e cervello umano nei due scienziati sia possibile solo a partire da una serie di apriori teorici tendenzialmente riduzionistici che vedono la macchina come modello interpretativo implicito dell’organico. La teoria della AI così come concepita dai suoi ideatori si tradurrebbe in una *reificazione* dell’intelligenza e dell’umano, che sarebbero pertanto conoscibili e riproducibili matematicamente.

Il riferimento a Kapp a tal proposito non è peregrino. Se nella modernità il rapporto tra organismo e macchina è stato declinato a partire da Descartes in termini riduzionistici di assimilazione del primo alla seconda¹³, Kapp nella sua teoria della proiezione legge la macchina come proiezione/imitazione del mondo organico, costituendo un potente strumento critico-teorico contro un approccio riduzionista e meccanicista.

2. Proiezione o disattivazione?

Il filosofo tedesco Ernst Kapp è ritenuto l’iniziatore della filosofia della tecnica. Nel 1877 diede alle stampe i *Grundlinien einer Philosophie der Technik*, lavoro per lo più ignorato dal dibattito filosofico dell’epoca, dove delinea la

¹³ Cfr. R. Descartes, *Discorso sul metodo*, in *Opere 1637-1649*, a cura di G. Belgioioso, Milano 2009, pp. 22-115. Sul ruolo del meccanicismo nello sviluppo delle odierne scienze cognitive si rimanda a J. P. Dupuy, *Alle origini delle scienze cognitive. La meccanizzazione della mente*, a cura di P. Heritier, Milano-Udine 2015. Sul meccanicismo cartesiano cfr. D. Garber, *Descartes, Mechanics, and the Mechanical Philosophy*, «Midwest Studies in Philosophy», 26 (1), Minnesota 2022, e S. Guidi, *L’angelo e la macchina. Sulla genesi della res cogitans cartesiana*, Roma 2018.

teoria della 'proiezione organica' (*Organprojektion*), suo principale contributo per una teoria della tecnica. La teoria della proiezione organica funge da criterio ermeneutico ed ontologico per individuare l'origine e la costituzione dell'ente tecnico a partire dalla sua dimensione prettamente antropologica, ponendo di fatto un'equazione tra tecnogenesi ed antropogenesi. Se tuttavia l'intento principale di Kapp nei *Grundlinien* è quello di capire l'uomo tramite la tecnica, qui si intende rovesciare il discorso andando oltre Kapp stesso ossia comprendere la tecnica e l'oggetto tecnico a partire dall'uomo.

Il concetto di *organon* delineato in tale contesto risente da un lato di una certa influenza del pragmatismo americano dell'epoca che Kapp conosceva da vicino – Frankliln aveva definito l'uomo come *tool making animal*¹⁴ – dall'altro di una concezione dell'ente artificiale di matrice aristotelica e della dialettica hegeliana, come avremo modo in seguito di chiarire. Sin dalla prefazione Kapp è chiaro sulla nozione di proiezione:

In primo luogo, viene dimostrato con irrefutabili dati di fatto che l'uomo conferisce inconsciamente la forma, la funzione e il rapporto normale della sua struttura corporea alle opere della sua mano e che solo dopo si accorga di questa relazione analogica di queste con se stesso. Questa realizzazione dei meccanismi a partire dalle forme organiche, la comprensione dell'organismo attraverso i dispositivi meccanici e innanzitutto la realizzazione del principio di proiezione organica come unica via possibile per il raggiungimento dei fini del fare umano è l'autentico contenuto di questo testo¹⁵.

Ben prima del celebre *Understanding Media* di McLuhann¹⁶, delle moderne teorie della *Material Engagement Theory*¹⁷ e della mente estesa, Kapp comprese che il rapporto di mediazione tecnica non è riducibile allo schema soggetto-oggetto-mondo; l'oggetto che è sempre estensione del corpo o di una facoltà, in quanto riproduzione inconscia di una operazione, retroagisce sull'organismo stesso e rappresenta, proprio in quanto sua oggettivazione, un paradigma di comprensione del corpo e delle facoltà umane. Ipotesi teorica di questo lavoro è che ciò sia valido tanto per il martello, quanto per le tecnologie dell'informazione e comunicazione (ICT) digitali e non, come il telegrafo elettromagnetico, esplicitamente citato nell'opera, e per la AI. Secondo Kapp infatti:

in relazione alla loro sede di generazione, essi [gli strumenti] offrono spiegazioni e comprensione dell'attività organica stessa, alla quale devono la loro origine come copia del modello. Questo deve essere riconosciuto come uno dei più importanti elementi sia per la teoria della conoscenza in generale che per lo sviluppo della coscienza di sé in particolare¹⁸.

¹⁴ E. Kapp, *Grundlinien einer Philosophie der Technik*, Hamburg 2015, p. 212.

¹⁵ Ivi, p. 4.

¹⁶ Cfr. M. McLuhann, *Understanding media. The extension of man*, London 2001.

¹⁷ Cfr. L. Malafouris, *How Things Shape the Mind: A Theory of Material Engagement*, Cambridge 2013.

¹⁸ E. Kapp, *Grundlinien einer Philosophie der Technik*, p. 172.

Del resto, la paleoantropologia ha dimostrato che sin dai primi reperti tecnologici della *pebble culture* l'ergonomia e il design rappresentano componenti fondamentali della produzione tecnica, volti a una crescente integrazione tra gesto, organo, interfaccia e strumento. Il design di questi primi enti tecnici per quel che riguarda la loro forma dimostra un progressivo affinarsi ergonomico e una sempre maggiore performatività, volta a potenziare la funzione organica in essi incorporata. Dal paleolitico al neolitico la lavorazione della pietra passa dalla semplice scheggiatura alla levigatura, compaiono materiali 'plastici' come la ceramica, nel Magdaleniano gli artefatti litici risultano ancora più finemente lavorati con la comparsa di aghi, frecce, lame sempre più affilate. Il processo tecnologico tende ad un'ottimizzazione della prestazione, a un aumento dell'ergonomicità delle interfacce e della potenza di azione dello strumento, questo è un tratto che accomuna la protoindustria litica al contemporaneo design industriale¹⁹. Forma e funzione sono dunque 'proiettati' dall'organo all'utensile. Ne reca una traccia l'etimologia dello stesso termine *organon*: in greco esso vuol dire essenzialmente strumento ed è poi passato nelle principali lingue moderne europee a significare la parte del corpo preposta a una determinata funzione. Lo stesso termine organico oggi indica l'opposto dell'artificio²⁰. Il rovesciamento dialettico è qui esplicito e conferma anche sul piano della semantica e della genealogia della lingua la continuità tra organismo e strumento descritte da Kapp: lo strumento, proiezione dell'organo, ha costituito il modello analogico di autorappresentazione dell'uomo delle sue prestazioni corporee ovvero il suo paradigma interpretativo. Successivamente il significato sedimentato del termine ha finito per perdere l'elemento di identificazione con l'ente artificiale per significare qualcosa che si contrappone proprio in quanto 'funzione organica' a ciò che è meccanico o artificiale. Questo ci testimonia in via di principio che l'analogia tra macchina e uomo è un qualcosa di assai più profondo e antico nella nostra cultura del meccanicismo moderno, o, nel campo della AI, dei tentativi di comprensione dell'attività cerebrale attraverso la sua modellizzazione artificiale. L'artificiale ha da sempre costituito il nostro doppio, lo specchio interpretativo dell'umano, una trasposizione che riguarda tanto l'oggettivazione fisica della funzione che l'ermeneutica del sé e la visione del mondo.

Secondo Kapp, tale trasposizione avviene in prima battuta inconsciamente, così come è inconscio il riferimento allo strumento nel termine organo. La forma anatomica non è il punto di partenza della creazione dell'artefatto, lo è la sua funzione o il fine dell'azione. Una volta che però l'oggetto è realizzato, il suo funzionamento rimanda all'organo modello. L'oggetto tecnico assume una connotazione ermeneutica nei riguardi dell'azione e della fisiologia del creatore.

¹⁹ Per la trattazione specifica della macchina come particolare oggetto tecnico cfr. Ivi., pp. 165 ss.

²⁰ È interessante notare come nell'industria alimentare il cibo organico (*organic food*) indichi proprio un tipo di alimenti non processati e coltivati secondo gli standard dell'agricoltura sostenibile, concimati con prodotti naturali, senza additivi e solventi. Rappresenterebbe quindi il modello alternativo rispetto alla agricoltura industriale, agli OGM, e alla agricoltura intensiva.

Ora, l'autocoscienza dell'uomo e il suo rispecchiamento nel mondo di apparati da lui costituito aumenterebbe col perfezionarsi delle tecnologie.

Ciò che è riprodotto non sarebbe solo uno specifico organo, ma anche il sistema nervoso e vascolare e in generale l'intera integrazione tra pensiero e azione per gradi crescenti di complessità. In questo contesto un ruolo di priorità, almeno cronologica, ce lo ha la mano. Scrive infatti Kapp:

L'uomo deve conoscere il proprio corpo, ciò significa semplicemente conoscere se stesso. Come si realizza questa conoscenza di sé con i mezzi molto concreti (*greifbar*) che sono gli utensili creati dalle sue stesse mani, questo è il nostro intero compito²¹.

La mano è a un tempo ciò che prende (*greifen*) e comprende (*begreifen*), crea lo strumento e il fine a cui l'utensile si deve conformare per la sua stessa utilizzabilità, determinando anche il concetto (*Begriff*) di questo come sua forma e finalità. D'altra parte, l'azione creativa della mano e del corpo in generale determina anche di ritorno la conoscenza dell'uomo stesso attraverso la mediazione tecnica²². La mano in generale costituisce il paradigma interpretativo stesso dell'agire, ne è testimonianza il termine tedesco *Handlung* (azione, atto) che conserva la parola *Hand* (mano) nella sua radice.

La conformità tra mano e utensile nasconde tuttavia un differenziale di potenza; la mano crea l'oggetto proprio a partire dalla presa di coscienza di un differenziale di potenza tra le nostre capacità corporee e gli oggetti realizzabili. Il lite scheggiato ha una potenza di taglio sicuramente superiore della mano, la macchina a vapore produce molti più kilowatt del metabolismo umano, e l'AI possiede una capacità di calcolo infinitamente maggiore rispetto a qualsiasi cervello umano. Ciò che tuttavia rimane a fondamento di queste prestazioni potenziate è il comune riferimento a una capacità umana e a una struttura organica che costituisce il modello di riferimento. La capacità prestativa aumenta, come sottolinea Fabio Grigenti²³, man mano che la tecnologia acquisisce nuove conoscenze e sviluppa apparati sempre più complessi. Scrive Grigenti:

Nel suo complesso il dominio dell'artificiale reca in sé la traccia evidente di una 'misura antropologica', la quale tradisce l'origine eminentemente umana – e perciò non naturale né divina – di tutti quegli ausili nei quali la nostra specie ha inconsapevolmente riprodotto, estendendole, particolari capacità già disponibili nel corpo vivente²⁴.

L'ente tecnologico è quindi espressione inconscia di ciò che è già a priori inscritto come dotazione specifica dell'uomo, ma non direttamente realizzabile attraverso la mera mediazione organica.

²¹ E. Kapp, *Grundlinien einer Philosophie der Technik*, p. 153.

²² Il nesso inscindibile tra conoscenza e manualità era stato già preconizzato da Anassagora di Clazomene, come ci testimonia il *De partibus animalium* di Aristotele in cui asseriva che l'uomo è il più intelligente degli altri animali perché ha la mano. Cfr. Aristotele, *Parti degli animali*, a cura di M. Vegetti, «Opere», vol. 5, Roma-Bari 1990, p. 127 (IV, 10, 687 a8-b5).

²³ F. Grigenti, *Le macchine e il pensiero*, Napoli, 2021.

²⁴ Ivi.

Pertanto, esso costituisce la via empirica privilegiata per comprendere lo sviluppo dell'intera cultura umana²⁵. L'uomo, secondo Kapp, crea un mondo esterno artificiale, nato da lui, che è espressione del suo interno distinto dalla natura tellurica e cosmica²⁶. L'uomo costituirebbe quindi come mondo proprio un mondo di apparati (*Apparatenwelt*) espressione del suo interno e in cui si riconoscerebbe²⁷.

Ciò è tanto più vero quanto la relazione tra organo ed oggetto tecnico è implicita. A tal proposito, Kapp porta l'esempio della relazione tra motore a vapore ed organi di trasmissione nella locomotiva e lo sviluppo di una rete infrastrutturale ferroviaria che ricalcherebbe la relazione tra cuore e sistema circolatorio. Tali innovazioni non avrebbero un esplicito legame con il nostro metabolismo o sistema circolatorio nella loro progettazione, bensì inconscio. Ciò permetterebbe un successivo riconoscimento, previa una alienazione strumentale inconscia, in cui l'uomo da un lato si riconosce 'nello' e 'a partire dallo' artefatto, dall'altro estende la sua impronta originaria sulla natura intesa in senso marxiano come terreno della lavorazione e trasposizione umana, come radura del dispiegamento della sua potenza o 'spirito oggettivo'. La proiezione non dà necessariamente vita a un riconoscimento diretto tra forma dell'organo e la sua riproduzione, ma può essere anche indiretta come nel caso delle macchine automatiche, in cui anzi l'organo è esonerato dallo svolgere la funzione. Qui entrerebbe in gioco un aspetto della mediazione tecnica che è stato evidenziato circa cinquant'anni dopo il lavoro di Kapp dal testo di Paul Alsberg *Das Menschheitsrätsel*²⁸ con il concetto di *Körperausschaltung*. Nel caso della macchina la funzione organica è in prima istanza inconsciamente trasposta, in secondo luogo è l'organo stesso, o in generale il corpo e la prestazione a essere disattivate o esonerate. Proiezione e disattivazione sono due facce della stessa medaglia, agiscono complementariamente nella macchina, ogni potenziamento ha come correlativo negativo una potenziale disattivazione. Vedremo successivamente cosa questo significhi nel caso della AI.

Ciò che nel testo di Kapp continuava a distinguere la macchina e il vivente è l'irriducibilità di quest'ultimo all'automatismo. Per comprendere ciò che intende il filosofo tedesco è utile riferirci alla caratterizzazione aristotelica di

²⁵ E. Kapp, *Grundlinien einer Philosophie der Technik*, p. 168.

²⁶ Ivi. p.171.

²⁷ Il termine *Apparatenwelt* è utilizzato da G. Anders ne *L'uomo è Antiquato, vol 1. Considerazioni sull'anima nell'epoca della seconda rivoluzione industriale*, tr. it. L. Dallapiccola, Torino 2007. C'è da notare che il rapporto tra uomo e mondo di apparati non è sempre di rispecchiamento, ma può trasformarsi, come del resto l'intera opera di Anders testimonia, in un rapporto di soggezione in cui l'uomo diviene prodotto dei suoi prodotti, rispetto a cui si troverebbe in una situazione di 'antiquatezza'. Da qui la 'vergogna prometeica', generata dal non essere all'altezza dei propri prodotti frutto del 'dislivello prometeico', ossia della differenza in termini di forza e potenza tra le capacità umane, fisiche, cognitive ed immaginative e i suoi prodotti tecnici. È interessante notare come la teoria della proiezione di Kapp sia armonizzabile con la teoria del dislivello andersiana, che ne costituirebbe il rovesciamento dialettico negativo.

²⁸ Cfr. P. Alsberg, *Lenigma dell'umano. Per una soluzione biologica*, tr. it. E. Nardelli, Roma 2021.

ente tecnico. Nel celebre passo della *Fisica B1*, Aristotele distingue l'ente tecnico dall'ente naturale a partire dall'eteronomia del movimento; l'ente tecnico non ha un impulso autonomo al movimento, al di là di quelle che possono essere le caratteristiche della sua struttura materiale (*scilicet*: naturale). Le realtà che sono frutto di *poiesis*, ovvero i prodotti della *techne*, non hanno in sé il principio di produzione, non sono in un certo senso cause di loro stesse. Ora, come è noto, la causa nell'orizzonte speculativo aristotelico è sempre quadruplici: materiale, formale, motrice e finale. Mentre nell'ente naturale tutte e quattro le cause coincidono con l'individuo stesso, per l'ente tecnico esse sono poste al di fuori. Ecco che entra in gioco lo *Architecton* che pone nell'ente tecnico la causa formale, la motrice e soprattutto la finale, mentre la materiale è in qualche modo presente direttamente o indirettamente nella *physis*. Se il comportamento orientato a uno scopo è quindi innato nell'animale, diremmo in termini fenomenologici che il vivente ha sempre un tipo di intenzionalità – cosa che sembra concedere anche Aristotele – nell'ente tecnico invece il fine è posto dallo *Architecton* o come scrive Kapp dallo *Hegemonikon*. È utile qui far riferimento direttamente a un passo delle *Grundlinien*:

Lo *Hegemonikon* della macchina non appartiene ad essa, non risiede in essa. Il fuochista e il macchinista di una locomotiva siedono su di essa e la dirigono, come il cavaliere il suo destriero. Lo *Hegemonikon* dell'organismo vivente, intelligenza e volontà appartengono ad esso, si sviluppano con esso, stesso costituiscono una sua parte fondamentale integrata²⁹.

Aggiunge poi Kapp in un passo assai rilevante per la nostra trattazione:

La visione del mondo meccanicistica e degradante della macchinizzazione dell'umano così come dell'umanizzazione della macchina è stata costruita con la suddetta argomentazione. L'assunto di Helmholtz secondo cui il concetto di lavoro della macchina sia stato preso dal paragone con l'uomo, nasconde l'immediata conseguenza per cui anche la macchina stessa, quando deve sostituire il lavoro dell'uomo – ovvero in modo corrispondente all'organismo di cui deve sostituire il lavoro – debba essere fatta in maniera corrispondente. La loro efficienza, o meglio la loro utilità, è direttamente in relazione all'uomo che ne ha necessità e allo scopo per cui determinati organi potrebbero operare anche senza l'assistenza delle macchine. La forma dell'organo appare più o meno chiaramente nei singoli utensili insieme alle prestazioni³⁰.

Tale passo è interessante e ci permette di far luce su tre aspetti fondamentali della teoria di Kapp. In primis Kapp legge differenza tra ente tecnico ed ente naturale secondo categorie aristoteliche, da cui ne consegue che la causa finale dell'ente tecnico non è nulla di tecnico in via di principio, ma appartiene alla *Lebenswelt* umana in quanto lavoro inserito in un sistema di bisogni, mentre la causa formale è più o meno direttamente un organo o una sua funzione. Inoltre, Kapp con straordinaria lucidità comprende come la visione meccanicistica

²⁹ E. Kapp, *Grundlinien einer Philosophie der Technik*, p. 132.

³⁰ Ivi.

e riduttivistica tipica della modernità e di un certo pragmatismo americano, si basa su un'aporia logica: dapprima essa interpreta e concepisce la macchina come *analogon* della azione umana (macchina e organismo funzionano allo stesso modo), in secondo luogo interpreta proprio sulla scorta di tale eguaglianza l'uomo come macchina, *de facto* meccanizzando l'umano e umanizzando la macchina. Dal passo emerge, inoltre, che l'utilità della macchina è sempre e solo in relazione all'uomo che ne ha necessità; da questo punto di vista l'impostazione di Kapp è decisamente umanistica.

In sintesi, il vivente nel suo crescere, nella sua intelligenza e volontà ha in sé il principio del suo movimento, è appunto 'auto-nomo' ha in sé la sua legge, la macchina la ha in altro; ciò fa però sorgere l'interrogativo circa le conseguenze ontologiche, etiche ed epistemologiche dal momento in cui una macchina possa, come nel caso della AI, raggiungere un certo livello di autonomia, e, in seconda battuta, se sia capace di un livello di *agency*, intenzionalità ed intelligenza come quella umana. Su questo punto le posizioni nel dibattito attuale sono varie e per certi versi opposte ed inconciliabili, fortemente polarizzate.

In un certo senso, Kapp aveva già assistito all'avvento di tecnologie superiori la cui filiazione dall'organo è indiretta e inconscia. Esempio di ciò sono la macchina a vapore e il telegrafo, tanto poco apparentemente morfologicamente connesse a un organo, da mettere in dubbio l'intero impianto teorico di Kapp. Tuttavia, a ben vedere, tali tecnologie interessano proiezioni organiche superiori; secondo il filosofo tedesco una tecnologia come il cavo elettrico, che ha permesso lo sviluppo e la diffusione del telegrafo e delle reti elettriche, ha come modello organico di riferimento il sistema nervoso. Ancora, la locomotiva ha finito per proiettare la vitalità cinetica dell'intero organismo in un sistema complesso di macchine e regolazioni. Gli attuali modelli di AI, reti neurali e lo *internet of things* assumerebbero e proietterebbero non solo la nostra intelligenza, ma una vastità di funzioni cognitive, sociali e organizzative diversificate come la comunicazione, l'immaginazione, il calcolo, la guida di autovetture e addirittura la società stessa con le sue gerarchie nei sistemi domotici complessi³¹. La cosa interessante da notare è che così come le funzioni fisiologiche e cognitive inferiori fondano quelle superiori, anche nelle tecnologie digitali della comunicazione il cavo elettrico come canale primario di conduzione dell'impulso ha un ruolo fondante, finanche nell'epoca della 'presunta' smaterializzazione dell'hardware nel cloud, dove i server continuano ad avere un ruolo primario per quanto dislocati³², e nelle intelligenze artificiali definite *Intelligent agent* che operano attraverso agenti inferiori disposti gerarchicamente.

³¹ Per una ricognizione su come gli algoritmi e la AI producono intelligenza sociale vedi anche E. Esposito, *Comunicazione artificiale. Come gli algoritmi producono intelligenza sociale*, Milano 2022.

³² Che la smaterializzazione dell'hardware e dei suoi consumi sia solamente un fenomeno apparente lo testimonia il report di Lean ICT, *Towards digital sobriety*, in cui si stima che secondo uno studio dell'università del Massachusetts Amherst, negli Stati Uniti, sviluppare una singola intelligenza artificiale determina l'emissione di 284 tonnellate di anidride carbonica e nel 2008, computer, dispositivi elettronici e infrastrutture digitali hanno contribuito per il 2% alle emis-

Ad esser 'proiettato', e in una certa misura esonerato, nel 'sistema tecnico' sarebbe quindi non solo il lavoro fisico umano, ma l'intero apparato di funzioni cognitive, fisiologiche e finanche sociali, essendo l'umano nella sua totalità il modello di riferimento di ogni produzione. In questo contesto potrebbe spiegarsi una tendenza che anima la AI generale sin dalle sue prime formulazioni e realizzazioni a ricreare macchine complesse in grado di riprodurre sempre più fedelmente prestazioni umane, esteriorizzando la totalità delle nostre facoltà e finanche il nostro vivere sociale. Secondo questo schema, sarebbe contestualizzabile anche la cibernetica come progetto di proiezione dei sistemi di organizzazione del vivente nei sistemi artificiali³³.

Il *telos* insito nella proiezione organica sarebbe quindi il continuo perfezionamento meccanico, che dal primo lite scheggiato condurrebbe alla 'macchina modello' (*Mustermachine*) che Kapp connota come strumento e tipo di apparato fisico che dovrebbe servire alla comprensione e alla riproduzione dell'azione reciproca che intercorre tra forze naturali e processi vitali all'interno dell'organismo³⁴.

La macchina modello sarebbe, pertanto, un ente artificiale che ci permetta di comprendere come modello oggettivato *in toto* i processi vitali e cognitivi all'interno del nostro organismo; ora se l'intelligenza, come insegna Nietzsche è l'ultimo e più imperfetto prodotto del corpo, è ovvio inferire che la perfetta macchina modello sia quella in grado di replicare la nostra volontà, intelligenza, immaginazione e creatività e i processi fisiologici e cognitivi sottesi. Il fine di questa macchina ideale non sarebbe altro quindi che la macchina organica. In essa il processo di regolazione, stabilizzazione e trasformazione dell'energia che

sioni globali di CO₂; nel 2020 al 3,7% e si prevede che toccheranno l'8,5% nel 2025, pari alle emissioni di tutti i veicoli leggeri in circolazione.

³³ La cibernetica nella formulazione dei suoi fondatori N. Wiener e W. R. Ashby focalizzava l'attenzione sull'autorganizzazione biologica e sulla analogia tra sistemi sia naturali che artificiali. L'utilizzo di strumenti creati per studiare il funzionamento di macchine costruite dall'uomo può essere esteso per spiegare alcuni degli attributi fondamentali degli esseri viventi. Questi attributi includono la tendenza al mantenimento della stabilità interna, la capacità di apprendere (cioè di modificare il proprio comportamento in base all'esperienza) e l'apparente scopo delle azioni. Questo approccio può essere diviso in tre nuclei principali di sviluppo. Il primo ha inizio negli anni '40 con i contributi di Ashby e riguarda l'estensione di questo metodo di analisi alle funzioni cognitive del sistema nervoso centrale e alle nuove macchine calcolatrici digitali, considerate unitariamente dal punto di vista del concetto di computazione. Il secondo nucleo, sviluppato tra il 1943 e il 1945 con i contributi di John von Neumann, Warren McCulloch e Walter Pitts, riguarda l'estensione dei concetti cibernetici alle funzioni cognitive del sistema nervoso centrale e alle nuove macchine calcolatrici digitali, considerate in modo unitario dal punto di vista del concetto di computazione. Il terzo nucleo, sviluppato durante le Conferenze Macy sulla Cibernetica tenute tra il 1946 e il 1953, riguarda l'applicazione dei concetti cibernetici allo studio delle scienze sociali. Wiener stesso, tuttavia, non contribuì direttamente a questa estensione e si dichiarò scettico sulla sua possibilità. Saranno studiosi provenienti dal campo delle scienze sociali, come Margaret Mead e Gregory Bateson, a promuovere e analizzare l'uso dei concetti cibernetici in questo contesto. Per i lineamenti fondamentali del pensiero di Wiener si rimanda a N. Wiener, *Introduzione alla cibernetica. Uso umano degli esseri umani*, tr. it. D. Persiani, Torino 2012. Sulla storia della cibernetica si veda S. J. Heims, *I cibernetici. Un gruppo e un'idea*, Roma 1994.

³⁴ E. Kapp, *Grundlinien einer Philosophie der Technik*, p. 133.

mantiene in equilibrio il sistema si integra con il controllo delle funzioni superiori. In questo senso, l'organismo è il sistema cibernetico perfetto. Secondo Kapp, dunque da un lato attraverso la macchina prendiamo sempre più coscienza della nostra dimensione e potenza organica, dall'altro la macchina nel suo disegno di potenziamento del vivente ha comunque nell'organico già segnato il binario dei suoi possibili sviluppi. Le macchine sono un'evoluzione in senso artificiale di ciò che è già insito nell'organico.

Da questo punto di vista, la macchina che più si avvicina alla definizione di macchina modello è proprio l'intelligenza artificiale generale. È giunto quindi il momento di vedere se l'argomentazione kappiana può dissodare un terreno epistemologico fecondo per la comprensione della AI. È quindi necessario prendere brevemente in esame la storia dell'intelligenza artificiale o almeno l'origine teorica del parallelismo mente-cervello nell'opera di Turing e von Neumann.

3. AI come macchina ideale

La storia dell'progetto dell'intelligenza artificiale è lunga e stratificata, secondo il matematico Martin Davis vedrebbe la sua origine nel *Calculemus!* di Leibniz³⁵, rimasto sorpreso dalla semplicità della numerazione binaria e dalla possibilità di ridurre la complessità dell'universo, incluso il ragionamento, ad un unico calcolo simbolico, interessa poi la storia della logica e vede come momento decisivo lo sviluppo delle macchine computazionali nell'800 ad opera di Babbage, e delle architetture di Turing e von Neumann. Intorno al 1840 Ada Lovelace preconizzò la possibilità di una macchina computatrice simbolica basata sulla logica³⁶; il tipo di macchina che aveva in mente era lo Analytic Engine progettato da Charles Babbage nel 1834 ma mai costruito. Lovelace riconobbe l'universalità del motore analitico e la sua potenzialmente infinita capacità di elaborazione dei simboli uguale a un computer digitale, prevedendone addirittura la possibilità per questa macchina di comporre elaborati scientifici e musicali di qualsiasi complessità, ma si fermò a descrivere alcuni elementi base della programmazione moderna senza dire di preciso come tale macchina avrebbe potuto realizzare quanto previsto.

La prima realizzazione di un computer elettronico programmabile si deve, come è noto ad Alan Turing un secolo dopo, che mise le sue conoscenze

³⁵ M. Davis, *Il calcolatore universale. Da Leibniz a Turing*, tr. it. G. Rigamonti e A. La Rana, Milano 2012.

³⁶ La matematica inglese Ada Lovelace figlia di Lord Byron, che è passata alla storia come la prima programmatrice per la sua partecipazione alla creazione della macchina analitica di Babbage, pubblicò nel 1843 un articolo contenente una traduzione di un lavoro del matematico italiano Luigi Federico Menambrea sulla macchina analitica di Babbage cui aggiunse una serie di note. Il testo fu pubblicato su *The Ladies Diary* e *Scientific Memoirs* di Taylor sotto le iniziali A. A. L.; in questo contesto Ada Lovelace intravedeva la possibilità di rendere la macchina analitica programmabile, prefigurando il concetto di intelligenza artificiale, restando tuttavia scettica sulla possibilità che tale macchina potesse divenire pensante come un essere umano.

scientifiche e logiche al servizio del Regno Unito durante la Seconda Guerra Mondiale. La prima Macchina basata sul modello di Turing³⁷, denominata *Colossus*, era un calcolatore in grado di decifrare i codici dei nazisti sviluppati dalla cifratrice Lorenz SZ 40/42 per proteggere le comunicazioni tra Hitler, i capi di stato maggiore del regime e gli alleati. *Colossus* fu progettato da Max Newmann sul modello della TM universale, realizzato da Tommy Flowers e installato nel famoso Bletchley Park sede delle operazioni di decodifica cui anche Turing prese parte.

Il progetto della TM è però di circa un decennio precedente, precisamente del 1936. Turing concepì la sua macchina ideale in analogia alle operazioni di calcolo umano nel suo *On computable Numbers, with an Application to the Entscheidungsproblem*³⁸, scritto che aveva come scopo fondamentale la confutazione dello *Entscheidungsproblem* di Hilbert³⁹. Il calcolatore umano lì descritto opera su un nastro lineare suddiviso in caselle – ad esempio un foglio – in cui scrive un solo simbolo, può utilizzare un numero limitato di caselle comprese nel suo spazio visivo, 'ricorda' i passi precedenti del calcolo per passare agli stati successivi e ha uno *stato mentale* come traccia presente nella memoria dai precedenti livelli del calcolo, tali stati mentali sono limitati dalla capacità di memoria del calcolatore. La lettura e la memoria sono sempre implicate nella lettura e nella manipolazione di simboli. Ora, secondo Turing sarebbe possibile in via di principio, replicare, e quindi esteriorizzare, le operazioni intelligenti del cervello umano in un programma di calcolo soggetto alle restrizioni di cui sopra. Tale programma è detto algoritmo di calcolo che riporterebbe la seguente istruzione:

Ac: a ogni istante del calcolo, ESEGUI l'azione determinata in modo UNIVOCO da ciò che OSSERVI sul nastro e da ciò che RICORDI dei passi precedenti⁴⁰.

Questa semplice istruzione corrisponde alla formalizzazione del pensiero umano in termini di *funzione* di esecuzione e predispone deterministicamente che in ogni istante del processo di pensiero una sola azione possa essere compiuta. Da qui Turing assume che sia possibile in via di principio da un lato ridurre ogni operazione del cervello umano a funzione, dall'altro creare una macchina in grado di simulare le operazioni del calcolatore umano. La TM è appunto questo calcolatore ideale che ha in sé esteriorizzata la funzione del calcolatore umano

³⁷ *Colossus* è in realtà stato anticipato dalla macchina Atansoff-Berry Computer, che però non era programmabile, la realizzazione di ENIAC, altro esempio di calcolatore digitale, è invece successiva. Il rigoroso e duraturo segreto imposto dal Regno Unito non permise a Colossus di vedere riconosciuto il suo primato, se non in tempi recenti.

³⁸ A. Turing, *On computable Numbers, with an Application to the Entscheidungsproblem*, «Proceedings of the London Mathematical Society», London 1936, pp. 230-265.

³⁹ In matematica ed informatica lo *Entscheidungsproblem* è il secondo problema di Hilbert posto nel 1928 da Hilbert e Ackermann. Il problema domanda di esibire una procedura meccanica (algoritmo) se una affermazione è un teorema della logica di primo ordine, ovvero se è derivato in un sistema di deduzione poggiante su soli assiomi di uguaglianza.

⁴⁰ E. Carli, F. Grigenti, *Mente, cervello, intelligenza artificiale*, Milano-Torino 2019, p. 103.

senza le limitazioni che tale macchina organica inevitabilmente si porta dietro. In particolare, non c'è un estremo superiore alla sua capacità di memoria, in quanto il nastro della macchina è indefinitamente espandibile, non si impone un estremo superiore alla durata dei processi di calcolo, infine, la macchina si assume funzioni sempre perfettamente e sia esente da errori di computazione. La TM è quindi definita da un insieme di regole che definiscono il comportamento della macchina su un nastro di input-output. Il nastro è diviso in celle ed è di una lunghezza ideale infinita, una testina legge i simboli spostandosi sul nastro. La TM opera essenzialmente sostituendo sul nastro due valori A e B (0 e 1), ma in via di principio è possibile ipotizzare, attraverso opportune operazioni di codifica, l'esistenza di una macchina capace di effettuare tutte le sostituzioni possibili dall'insieme S delle macchine di sostituzione. Una tale macchina è chiamata la Macchina Universale di Turing (UTM), macchina astratta che non eseguirà solo un algoritmo, ma ogni algoritmo possibile, e quindi ogni altra TM. La UTM sarebbe quindi un ente artificiale potenzialmente in grado di simulare un numero n di processi logici e aspetti del pensiero umano; pertanto, sorgerebbe spontanea la domanda se in via di principio sia possibile concepire una TM in grado di imitare tutti gli aspetti del pensiero a patto che questi possano essere ridotti a funzione.

Questo meccanismo è l'archetipo di ogni computer e intelligenza artificiale e, come si vede, è stato realizzato da Turing con l'intento preciso, non inconscio questa volta, di exteriorizzare una o più funzioni del cervello, a partire da un determinato a priori ermeneutico di come esso 'pensa'. Turing concepisce il pensiero umano essenzialmente come calcolo, per cui è in via di principio sempre possibile, come riteneva prima di lui Hilbert, esprimere il pensiero in simboli codificati. Nel caso della TM la teoria dell'esteriorizzazione di Kapp appare calzante, la funzione qui exteriorizzata sarebbe la capacità di calcolo a cui Turing, meccanicisticamente, riduce il pensiero umano. La TM exteriorizza, potenzia, ma anche esonera il calcolatore umano superandolo di gran lunga in velocità, performatività, potenza e memoria di esecuzione. Nel suo *Computer Machinery and Intelligence* del 1950 apparso sulla rivista *Mind*⁴¹, Turing si spinge ancora oltre ponendo esplicitamente la domanda se una macchina possa pensare, ovvero se possa avere una *agency*. Come specificato precedentemente, Turing era ben cosciente delle aporie che una tale domanda si porta dietro, in primis una definizione dei termini macchina e pensare. Il matematico inglese giunse quindi nell'articolo a elaborare un test, noto come *Turing test*, secondo cui per determinare se una macchina è realmente intelligente, essa deve esibire un comportamento tale da essere in grado di ingannare il suo interlocutore circa la sua vera natura. Il Test di Turing è stato più volte riformulato nel corso degli anni, sia perché la sua formulazione originale appare imprecisa sia perché lo sviluppo di nuove macchine intelligenti in grado di simulare troppo facilmente il comportamento umano hanno posto problemi nuovi. Così come note sono

⁴¹ A. Turing, *Computer Machinery and Intelligence*, «Mind», 59, Oxford 1950.

le obiezioni a tale test, tra cui la più celebre è senz'altro la 'stanza cinese' di John Searle. Il filosofo americano presentò l'argomentazione nell'articolo *Minds, Brains and Programs*, pubblicato nel 1980 dalla rivista scientifica *The Behavioral and Brain Sciences*⁴².

Il filosofo americano affermava l'impossibilità di inferire un comportamento intelligente da parte di una macchina dalla semplice capacità di ingannare l'interlocutore. Alla base di tale contro-test c'è la convinzione che la sintassi non sia condizione sufficiente per determinare la struttura semantica di un enunciato. L'argomentazione fu presentata dal filosofo americano nel 1980 in un articolo dal titolo *Minds, Brains and Programs* nella rivista *The Behavioural Brain and Science*⁴³. Questo argomento, come ci testimonia tra gli altri Floridi⁴⁴, è tutt'oggi un'obiezione contro l'intelligenza artificiale forte. In questo contesto Searle contesta a Turing la possibilità di assimilare gli stati cognitivi umani a un programma preimpostato (algoritmo). Essenzialmente l'argomentazione di Searle parte dal concetto di intenzionalità e si configura a un tempo come critica al test di Turing, alla sua implicita filosofia, e alla possibilità di realizzare una *strong AI*. Le Argomentazioni sono le seguenti:

1) L'intenzionalità negli esseri umani e negli animali è il risultato delle caratteristiche causali del cervello. Certi processi cerebrali sono sufficienti per l'intenzionalità.

2) Istanziare un programma informatico non è sufficiente per raggiungere l'intenzionalità. L'articolo sostiene questa tesi mostrando come un agente umano potrebbe avere il programma ma ancora mancare dell'intenzionalità rilevante.

Da cui ne consegue:

a) L'esplicazione di come il cervello produca l'intenzionalità non può essere attribuita unicamente all'istanziamento di un programma informatico.

b) Qualsiasi meccanismo capace di produrre intenzionalità deve avere poteri causali pari a quelli del cervello.

c) Qualsiasi tentativo di creare artificialmente l'intenzionalità (*strong AI*) richiederebbe di duplicare i poteri causali del cervello umano, non solo progettare programmi.

Secondo l'argomento di Searle, solo le macchine che hanno capacità causalmente equivalenti a quelle del cervello umano possono veramente pensare e avere intenzionalità. La *strong AI*, che si focalizza esclusivamente sui programmi, non può fornire una comprensione del pensiero poiché esso dipende dalle

⁴² J. Searle, *Minds, Brains and Programs*, «Behavioral and Brain Sciences» Volume 3, Issue 3, September 1980, pp. 417-424.

⁴³ J. R. Searle, *Minds, brains, and programs.*, pp. 417-457.

⁴⁴ Cfr. L. Floridi, *La Quarta Rivoluzione*, Milano 2017, pp. 256 ss.

capacità causali interne della macchina o del cervello, e non semplicemente dal programma in sé. In altre parole, il pensiero e l'intenzionalità richiedono più di un semplice esecutore di programmi, ma implicano un'elaborazione causale simile a quella presente nel cervello umano⁴⁵. In termini di teoria della proiezione, permarrebbe uno scarto tra la dimensione organica e la sua exteriorizzazione, colmabile solo da una esatta riproduzione dei meccanismi causali del cervello.

Tale complessità non era considerata da Turing, secondo cui il comportamento della TM ideale sarebbe assimilabile a quello del cervello umano, previa riduzione dell'attività cerebrale a funzione. Questa ipotesi prende il nome di computazionismo, una forma particolare di riduttivismo in cui i processi mentali sono riducibili a processi di calcolo, quindi datificabili, operanti tramite simboli; su questa base i processi mentali sono equivalenti a quelli operati da un computer.

Sebbene la TM exteriorizzi delle funzioni del nostro cervello, tuttavia, secondo Searle i procedimenti e il funzionamento del cervello organico e di quello artificiale non sono assolutamente sovrapponibili. Per il filosofo la mente umana si distingue da qualsiasi altra perché dotata di intenzionalità, termine assai stratificato in storia della filosofia che rimanda alla riflessione medioevale di Avicenna e Abelardo e alla fenomenologia di Husserl. *Intentio* nella filosofia medioevale indica l'esser sempre riferito della coscienza a un qualcosa, e coscienza è sempre 'coscienza di', il che ci riporta alla sua connotazione semantica. Il concetto è *sermo* (discorso) riferito a un'entità reale fuori di sé e tale riferimento implica sempre un rapporto di significazione tra enunciato e realtà significata⁴⁶. Similmente per Searle ogni azione e stato cognitivo implicano sempre un riferimento al significato del termine che media il rapporto con l'oggetto fuori di esso. L'intelligenza artificiale nel suo operare mediante simboli non è dotata di intenzionalità, essendo il suo comportamento basato sulla elaborazione di tali simboli in accordo a una regola sintattica (la funzione) e non sul riferimento o la significazione. Inoltre, Searle fonda l'intenzionalità dal punto di vista biologico: il differenziale tra le due intelligenze è basato essenzialmente sul fatto che l'intelligenza organica è sempre un prodotto biologico frutto di una evoluzione e di processi bio-chimici che non possono essere in via di principio ridotti a semplici impulsi elettici come sostengono i teorici della AI forte⁴⁷. Nel Turing test, pertanto, per quanto intelligenza artificiale e umana possano avere gli stessi risultati, differenti sono le procedure per ottenerli. Gli elementi che la AI manipola non comportano la comprensione delle istruzioni, i procedimenti non sono mediati dal significato ma dall'automazione. L'esteriorizzazione dei

⁴⁵ Cfr. J. R. Searle, *Minds, brains, and programs*.

⁴⁶ Per un ulteriore inquadramento della questione della intenzionalità nella filosofia medioevale e in particolare in Abelardo per il ruolo che gioca nella disputa sugli universali si rimanda a E. Gilson, *La filosofia del medioevo. Dalle origini patristiche alla fine del XIV secolo*, Firenze 1973, pp. 336 ss.

⁴⁷ Cfr. J. Searle., *La riscoperta della mente*, Torino 1994, p. 110.

processi cognitivi umani sarebbe in via di principio possibile nei risultati ma non nella sostanza. Le proprietà formali del programma non sono assimilabili alla intenzionalità e non hanno di per sé poteri causali, in quanto non si può separare la mente, intesa come programma e cervello biologico, a meno che non si riconosca l'indipendenza delle due strutture, cosa alquanto improbabile. Alla domanda quindi se le macchine possano pensare in senso forte, Searle risponde in maniera assai più netta di Turing affermando che esse potrebbero farlo se e solo se in grado di replicare un sostrato fisiologico bio-chimico e neurologico. Siamo naturalmente ancora lontani da una tale evenienza.

La controprova che Searle appronta al test di Turing ripropone le stesse condizioni di partenza del test in cui un umano interagisce con una macchina chiusa dentro una stanza. Searle si sostituisce alla macchina e immagina di interagire con un interlocutore cinese all'esterno. Supponiamo inoltre che il filosofo non abbia alcuna conoscenza del cinese e che nella stanza ci siano una serie di caratteri che Searle dovrà utilizzare per rispondere all'interlocutore. Ora il cinese non ha alcuna somiglianza semiotica ed etimologica con l'inglese, il filosofo, pertanto, non ha mezzi per formulare alcun enunciato, ma ha a disposizione solo simboli. Nella stanza c'è tuttavia un libro di istruzioni – l'equivalente del programma – in cui i caratteri cinesi sono associati tra loro secondo regole scritte in inglese. Il filosofo continuerà a non sapere assolutamente nulla del cinese, ma potrà rispondere alle domande poste dall'interlocutore grazie alle indicazioni in inglese del manuale. Tali indicazioni corrispondono al programma che permette al computer di mettere in relazione simboli formali con una serie di altri simboli formali, dando una risposta (output) a una domanda (input). Dal punto di vista puramente ipotetico questo permetterebbe a Searle di interfacciarsi con l'interlocutore cinese in grado direttamene proporzionale all'esattezza delle indicazioni del manuale. Più le indicazioni saranno compresse e precise, più il programma sarà complesso, maggiore sarà la capacità di Searle di interfacciarsi con il madrelingua al punto di essere considerato a sua volta tale. In questo contesto, è evidente come la conoscenza della lingua non sia dirimente quanto piuttosto la capacità di seguire le istruzioni fornite. Il rapporto con la lingua cinese sarà dunque sintattico e non semantico, non si collega a ciascun termine un significato, ma si sanno seguire adeguatamente le istruzioni fornite. Questo è quanto accade *mutatis mutandis* con l'AI quando questa tenta di imitare le operazioni del pensiero umano, essa agisce proceduralmente ma non comprende ciò che fa, così come Searle nel suo esperimento immaginario non conosce il cinese.

L'argomento di Searle poggia, come si è detto su una differenza di fondo dei due sistemi intelligenti: quello organico è fondato su processi biologici, neurologici e intenzionali, quello artificiale su un programma di calcolo. Se dovessimo, pertanto, riferire queste considerazioni alla questione della esteriorizzazione, ne risulterebbe che la proiezione nella TM dell'intelligenza umana riguarda solo e solamente la capacità computativa del cervello, ma non è assimilabile alla totalità della sua attività intenzionale, senza una riduzione

forzosa di quest'ultima alla prima. Inoltre, tanto per Kapp, come per Searle tra il modello organico e la sua proiezione tecnica permangono sempre a un tempo un legame di causazione e un irriducibile iato funzionale.

La convinzione che la AI fosse possibile furono rafforzate dal lavoro seminale di W. McCulloch e W. Pitts dal titolo *A logical calculus of the Ideas Immanent in Nervous Activity*⁴⁸. I due autori aggiunsero al lavoro di Turing la logica proposizionale di Bertrand Russell e la teoria delle sinapsi di Charles Sherrington, poiché entrambe descrivevano sistemi binari: vero e falso per la logica proposizionale e attività on/off delle cellule cerebrali. Questi due stati vero/falso e on/off furono poi assimilati ai singoli stati 0/1 della TM. Secondo Sherrington i neuroni avevano anche delle soglie di attivazione fisse, i connettori logici *e*, *o* e *non* servivano da porte logiche definite come reti neurali connesse tra loro per la rappresentazione di proposizioni complesse. Di conseguenza tutto ciò che poteva essere affermato secondo la logica proposizionale era rispecchiato in una rete neurale e computato da una TM. L'argomentazione dei due scienziati poggiava su una eccessiva e superficiale equiparazione di logica, computazione e neurofisiologia e sulla equiparazione tra linguaggio naturale e logica. Ciò garantiva un unico approccio teorico alla intelligenza umana e alla intelligenza artificiale, realizzando di fatto l'eguagliamento ontologico, funzionale e pragmatico tra le due con la mediazione della teoria della computazione di Turing. Il lavoro di W. McCulloch & W. Pitts fu fondamentale per lo sviluppo di quella che storicamente ha preso il nome della GOFAI⁴⁹.

Dopo Turing, un altro illustre matematico ha provato a fondare l'analogia tra computer e cervello proprio partendo dal funzionamento bio-chimico di quest'ultimo. Il matematico ungherese John von Neumann, che aveva, tra le altre cose⁵⁰, precedentemente lavorato con Hilbert al disegno di fondazione della matematica, nel suo *The Computer & the Brain*, in continuità con il progetto di Turing, istruì esplicitamente il parallelismo tra computer e cervello umano e quindi tra tecnologie informatiche e neuroscienze proprio a partire da una similitudine di funzionamento. Von Neumann si era interrogato sull'enigma dell'autorganizzazione dei sistemi e, proprio sulla scia del lavoro di McCulloch & Pitts, si era servito della loro impostazione teorica per spiegare l'evoluzione biologica e la riproduzione degli automi cellulari. Questi ultimi sarebbero sistemi di unità computazionali i cui cambiamenti dipendono dallo stato delle unità

⁴⁸ Cfr. W. McCulloch & W. Pitts, *A logical calculus of the Ideas Immanent in Nervous Activity*, «Bulletin of Mathematical Biophysics» Volume 5, Oxford 1943.

⁴⁹ Nella filosofia della AI, GOFAI ('Good old fashioned artificial intelligence') rappresenta l'approccio classico della AI simbolica, in contrasto con altre metodologie, come le reti neurali, la robotica situata, la AI simbolica limitata o AI neuro-simbolica. Il termine è stato coniato dal filosofo John Haugeland nel suo libro *Artificial Intelligence: The Very Idea*, Princeton 1985.

⁵⁰ Per una biografia sulla vita fuori dall'ordinario di J. von Neumann, che come è noto partecipò anche al Manhattan Project, elaborò la teoria dei giochi in economia, costruì modelli automatici di riproduzione cellulare (*von Neumann cellular automation*), si occupò di fluido dinamica e di metereologica, si rimanda a A. Bhattacharya, *The Man from the Future*, London 2022.

vicine e da semplici regole. Queste unità sono in grado di replicarne altre, gli errori di replicazione possono condurre all'evoluzione. Gli automi cellulari, come vedremo sono il modello di riferimento in cui è proiettato il funzionamento dei neuroni.

Il disegno teorico di von Neumann si basava su un'asserzione di principio, mutuata da McCulloch & Pitts, che qualsiasi scambio d'informazione analogica può essere tradotto digitalmente mediante l'uso della logica. Ciò implica in via di principio che l'architettura della cognizione umana può essere sovrapponibile a quello di una TM e tradotto in valori 0/1.

Applicando il concetto di universalità del calcolo, egli sosteneva che, per quanto le unità e le architetture di un cervello organico e di un computer possano apparire sotto ogni punto di vista differenti, è possibile affermare che una macchina è in grado in linea di principio di simulare, e quindi proiettare, i processi di elaborazione del cervello, ma non viceversa poiché il cervello non ha un programma memorizzato; il *logos* umano non è un programma. L'assunto è funzionale da un lato a salvare la specificità del pensiero, ma dall'altro a determinare l'analogia tra macchina e cervello. Von Neumann è autore di un particolare tipo di macchina, che aggiorna e implementa la Turing Machine come modello di riferimento di ogni computer. Il suo modello è l'anello di congiunzione tra il calcolatore universale di Turing e la cibernetica come scienza dell'organizzazione della relazione di sistemi complessi.

L'architettura della cosiddetta Macchina di von Neumann è basata sulla RAM (*Random Access Memory*) ed ha costituito il modello di computer più diffuso, almeno fino all'avvento dei computer a reti neurali e dei network computer⁵¹. In questa macchina l'esteriorizzazione dei processi e delle strutture cognitive e mnemoniche del cervello organico è stata coscientemente elaborata da von Neumann, in una straordinaria, quanto probabilmente involontaria, continuità rispetto alle strutture evidenziate da Kapp. I principi di questa tecnologia furono introdotti in un articolo del 1945 intitolato *First Draft of a Report on the EDVAC*. Il modello di von Neumann include un'unità di elaborazione centrale, definita processore, dove vengono effettuate le operazioni logico-aritmetiche, un'unità di memoria (RAM) nella quale sono memorizzati il programma e i dati, una memoria di massa, un contatore di programma e i canali di input e output⁵². Rispetto alla TM, la macchina di von Neumann introduce un sistema di calcolo a otto bit, inoltre se il nastro di memoria della prima macchina era sequenziale, per cui essa impiegava un tempo notevole per muovere il nastro avanti e indietro per recuperare le informazioni, la memoria ad accesso casuale (RAM) permette di recuperare i dati in qualsiasi momento. Un'ulteriore implementazione del modello di Turing consiste nel concetto di programma memorizzato sulla stessa

⁵¹ Da questo punto di vista occorre precisare che proprio l'avvento della cibernetica e del connessionismo, che per motivi spazio possiamo solo accennare in questo frangente costituisce un passaggio decisivo, poiché ha permesso la connessione sistemica di macchine di Turing tra loro e lo sviluppo a partire dagli anni '60 delle ANN (Artificial Neural Network).

⁵² Cfr. J. von Neumann, *Computer e cervello*, tr. it. P. Bartesaghi, Milano 2021, p. 18.

RAM su cui sono allocati i dati, ciò consente al computer di essere riprogrammato in funzione dei compiti e di poter operare secondo codici automodificanti, consentendo operazioni ricorsive. Questo permette al computer di svolgere una serie di operazioni diverse, consentendo, almeno in via di principio secondo il disegno teorico di von Neumann, una polifunzionalità simile al cervello umano che è il modello di riferimento, sia per quanto riguarda la memoria sia per i dispositivi di input ed output. Un'ulteriore innovazione introdotta dal matematico ungherese nel progetto EDVAC⁵³ è l'inserimento di un codice di operazione in ogni istruzione che specifichi l'operazione da eseguire e l'indirizzo in memoria dell'elemento su cui essa deve agire. Infine, l'ultima trovata di von Neumann, fondamentale per il tema di questo saggio, è la possibilità esplicita, preconizzata già come detto Ada Lovelace, di dotare le macchine di intelligenza. A questa esigenza risponde l'ormai classico saggio *Computer and Brain* del 1958. In queste pagine, a differenza di Turing, il cui lavoro del 1950 era ben noto a von Neumann, il matematico comprende bene che per poter rispondere alla domanda se una macchina possa o meno essere dotata di intelligenza, bisogna in via preliminare discutere se esista e in che termini tra essa e il cervello organico un parallelismo, almeno dal punto di vista funzionale. A ben vedere, ogni possibilità di identificare o realizzare una AI forte si gioca proprio su questa analogia fondamentale che almeno von Neumann ritiene in via di principio possibile. Ed è proprio nelle pagine di *Computer and Brain* che l'analogia tra cervello e macchina viene presa in esame, sulla scia del lavoro di Turing e soprattutto del lavoro di McCulloch & Pitts.

Ciò vuol dire istruire contemporaneamente il parallelismo tra informatica, psicologia e neuroscienze, che è di fatto il lascito più rilevante dell'opera di von Neumann, scritta peraltro poco prima che il cancro lo portasse via. Questo parallelismo ha alimentato la stagione della cibernetica di cui von Neumann si può considerare un fondatore⁵⁴ ed è ancora oggi alla base di tutte le teorie della AI, dalla GOFAI alla programmazione basata su regole, secondo cui la psicologia umana in generale è rappresentabile in inferenze se-allora⁵⁵, fino alle reti neurali artificiali. Tale parallelismo è anche alla base della possibilità di riconoscere l'esteriorizzazione e la proiezione come modello ermeneutico della AI.

Assunto fondamentale dell'argomentazione e fulcro teorico dell'analogia è la constatazione che i neuroni producono un output digitale sotto forma di impulso

⁵³ *L'Electronic Discrete Variable Automatic Calculator* (EDVAC) è uno dei primi calcolatori digitali della storia, uno dei a programma memorizzato basato sull'architettura di von Neumann, rispetto al suo predecessore l'ENIAC, era basato su un sistema di numerazione binario e non più decimale.

⁵⁴ Cfr. M. A. Boden, *L'intelligenza artificiale*, pp. 18-21.

⁵⁵ H. Simon ed A. Newell che insieme a J. C. Shaw realizzeranno *Logic Theorist*, il primo programma deliberatamente pensato per il ragionamento automatizzato, riterranno che lo schema se-allora è alla base della psicologia umana in generale. La loro creazione, presentata alla celebre e fondazionale *Dartmouth Summer Research Project on Artificial Intelligence*, sarà in grado di risolvere 38 dei primi 52 teoremi dei *Principia Mathematica* di Whitehead e Russell. Questo modello basato su regole non specifica in anticipo l'ordine dei passaggi come l'architettura di von Neumann e aggiorna i programmi basati sul controllo esecutivo.

elettrico. Un assone, similmente al sistema binario dei computeri digitali, si attiva o non si attiva, l'impulso passa o non passa. Tuttavia, l'analisi del segnale che passa nei dendriti al soma del neurone è di tipo analogico; ne risulta che, mentre una macchina di von Neumann è un oggetto totalmente digitale, il cervello umano è ibrido, analogico per alcune funzioni, digitale per altre. Come sottolinea Ray Kurzweil nella prefazione a *Computer and Brain*⁵⁶, questa impostazione portò allo sviluppo del connessionismo, un modello per la programmazione degli hardware e dei software in cui i sistemi si ispirano al funzionamento del neurone. Il connessionismo tenta di spiegare i processi mentali attraverso lo sviluppo di reti neurali artificiali. Tale prospettiva, delineata per la prima volta da David Rumelhart e David McClelland, si pone in antitesi al cognitivismo che sostiene l'integrale analogia tra mente e computer. Il connessionismo considera invece il computer un mero strumento di calcolo e sposta l'analogia funzionale sul cervello umano e le reti neurali biologiche. Secondo questo schema la mente deve la sua attività alle connessioni tra unità singole di calcolo – i neuroni – e non è scomponibile tra i singoli processi cognitivi. Come il comportamentismo, la prospettiva connessionista postula una connessione tra stimolo e risposta che fonda l'analogia tra processi organici e artificiali. Tuttavia, è evidente come qui la mente non sia considerata mero calcolo come nel simbolismo di Turing, e non sia basata su procedimenti solamente sintattici⁵⁷, ma risultato di processi di apprendimento⁵⁸ che coinvolgono stimoli ambientali esterni e memorizzazione di queste informazioni nei collegamenti tra neuroni. Il connessionismo è alla base dello sviluppo delle reti neurali artificiali, che grande fortuna hanno avuto nello sviluppo dell'intelligenza artificiale, del ANN e del *deep learning*. Ogni neurone artificiale prende nome di nodo, terminologia ancora oggi in uso, e deve il suo funzionamento e regolazione da una funzione di attivazione che stabilisce quando questo invia il segnale⁵⁹. Tali reti opererebbero in parallelo secondo lo schema del cervello umano e non in serie come la macchina di Turing.

A questo livello, il processo di proiezione di una funzione e di una struttura organica è operato scientemente dai connessionisti; ciò dimostra nei fatti come il paradigma della proiezione organica di Kapp possa dimostrarsi un criterio

⁵⁶ M.A. Boden, *L'intelligenza artificiale*, p. 23.

⁵⁷ Il tentativo più integrale di riduzione dei processi mentali a processi sintattici si deve alla teoria della mente computazionale di Jerry A. Fodor. Cfr. J. A. Fodor, *Mente e linguaggio*, a cura di F. Ferretti, Bari 2001.

⁵⁸ Un precursore di questa tendenza si ha in W. R. Ashby che nel suo *The Nervous System as Physical Machine: With Special Reference to the Origin of Adaptive Behavior*, «Mind, New Series» Vol. 56, No. 221. (Jan., 1947), pp. 44-59) teorizza la possibilità di rappresentare l'organizzazione dei neuroni e la loro coordinazione e il comportamento adattivo in termini fisico-matematici. In quel contesto il comportamento adattivo è descritto nel riequilibrio delle funzioni dopo un evento perturbante. Ashby interpreta il sistema nervoso come una macchina che tende all'equilibrio con il suo ambiente, o meglio che tale macchina sia rappresentata proprio dall'insieme delle interazioni tra sistema nervoso e ambiente.

⁵⁹ Per ulteriori approfondimenti circa questa teoria si rimanda a D. E. Rumelhart, J. L. McClelland and the PDP Research Group, *Parallel Distributed Processing: Explorations*, «Microstructure of Cognition» Volume 1: Foundations, Cambridge 1986.

epistemologicamente valido per comprendere l'ontogenesi delle intelligenze artificiali, sia per quel che riguarda l'Architettura di von Neumann che per le reti neurali artificiali⁶⁰. Dunque, sia von Neumann che il connessionismo, sebbene rifiutino l'equiparazione immediata di Turing tra mente e calcolo, ritengono che il computer possa simulare i processi di elaborazione del cervello, ma che non sia vero il contrario, poiché il cervello non è dotato di un programma memorizzato e, a causa della sua natura ibrida analogico-digitale, non è in grado di replicare la precisione e la velocità di tali macchine 'intelligenti'.

Occorre però segnalare un'altra differenza cruciale, su cui von Neumann appare sorvolare: i neuroni, a differenza della macchina, apprendono i modelli dai loro stessi input, codificati attraverso differenti concentrazioni di neurotrasmettitori, e dalla costruzione e distruzione di connessioni, come la teoria del darwinismo neurale di Edelman ha ampiamente dimostrato⁶¹; il cervello animale sarebbe dunque un'entità plastica.

Rispetto alla sua macchina poi, von Neumann osserva che la velocità di elaborazione dell'informazione di un neurone è estremamente bassa, ma il sistema compensa tale lentezza attraverso l'elaborazione in parallelo. Von Neumann arriva persino a stimare la capacità di memoria del cervello umano⁶², non cogliendo però il fatto che a differenza della RAM di un computer in cui i dati immagazzinati possono anche a distanza di anni essere richiamati e letti senza possibilità di errore, la nostra mente tende a perdere ricordi, pensieri ed esperienze, a dimenticarli o ad archivarli, in quanto la nostra coscienza è organizzata secondo una gerarchia di schemi in cui il gradino più basso è costituito dagli schemi senso-motori, i dettagli delle esperienze non vengono sempre memorizzati in toto. Inoltre, la nostra capacità decisionale in frazioni di secondo implica che i (presunti) algoritmi sequenziali alla base del funzionamento del nostro cervello non possono essere troppo complessi. Anche da una prospettiva meccanicistica come quella di von Neumann, l'equiparazione non è così immediata, almeno non nei modi di un integrale rispecchiamento.

⁶⁰ Il primo sistema connessionista è stato creato nel 1957 alla Cornell University da Frank Rosenblatt come software per un computer IBM 704.

⁶¹ Cfr. G. M. Edelman, *Darwinismo neurale. La teoria della selezione dei gruppi neuronali*, a cura di S. Ferraresi, Milano 2018. Edelman, una delle più forti voci critiche nel dibattito attuale circa la possibilità di una intelligenza artificiale forte, partendo dalla stessa prospettiva di von Neumann, lo studio dei processi neurologici, arriva a conclusioni diametralmente opposte. La teoria del darwinismo neuronale poggia sull'idea secondo cui le funzioni superiori del cervello siano il risultato di uno sviluppo filogenetico di una data specie e delle variazioni anatomiche e funzionali presenti alla nascita di ogni singolo organismo animale. Questo approccio che vede come in Searle e Kapp il primato della struttura biologica della mente è opposto al funzionalismo materialista di Turing e Putnam. Edelman ritiene che gli stati mentali non possano essere tradotti in un software. Edelman lavora su modelli di intelligenza artificiale debole che simulano determinate parti del cervello, essi non hanno l'ambizione di risolvere tutta l'attività cerebrale, ma solo alcune funzioni, inoltre in essi le scelte non sono programmate da un algoritmo come nella TM, ma conseguenza di addestramento (*machine learning*), simulando l'apprendimento dell'uomo.

⁶² J. von Neumann, *Computer e cervello*, pp. 118-122.

Cervello e computer risulterebbero, pertanto, paragonabili nei loro risultati ma non sovrapponibili dal punto di vista funzionale. Von Neumann conclude che un computer moderno potrebbe in via di principio replicare la stessa operazione cognitiva del cervello, ma seguendo percorsi differenti. Analogia quindi, ma non sovrapponibilità delle intelligenze. I meccanismi analogici del cervello possono essere replicati attraverso il calcolo digitale in qualsiasi livello di precisione desiderato, in quanto la precisione di un sistema analogico che agisce in parallelo è assai bassa, ma non è vero il contrario; ciò contravviene ad ogni principio di identità ($A=B$, ma $B \neq A$; dove A è una operazione cerebrale e B una operazione della macchina). In poche parole, se il funzionamento in parallelo del cervello umano può essere simulato attraverso l'enorme potenza di calcolo dei calcolatori, la potenza di calcolo di questi ultimi non può essere simulata da alcuna intelligenza organica, senza contare che i moderni supercomputer operano proprio in parallelo.

In conclusione, il paradigma della proiezione può essere applicabile sia alla macchina di Turing che alla architettura di von Neumann; entrambi i modelli di computer digitali nascono con il chiaro intento di replicare il funzionamento del cervello umano e di superarne le prestazioni. Tuttavia, i due modelli sono viziati da una *petitio principii* ossia l'equiparazione tra logica, informatica e neuroscienze. Parallelismo che non ci sentiamo di condividere perché, in quanto frutto epigonale del meccanicismo moderno, non rende ragione della base teorica, epistemologica e scientifica di tale equiparazione, ma la presuppone. Questo principio epistemologico è risultato essere frutto di una temperie culturale che ha attraversato tutto il Novecento interessando anche le scienze cognitive e la cibernetica, le teorie comportamentiste e connessioniste. Siamo ora pronti per trarre le conclusioni di questa nostra analisi.

4. Conclusioni: sulla possibilità della esteriorizzazione come modello teorico

L'intento della trattazione è applicare il principio di proiezione organica di Kapp alla genesi dell'intelligenza artificiale, nella sua prima formulazione teorica. Per far ciò, dopo aver introdotto la questione della AI nel dibattito attuale e dopo aver tratteggiato i lineamenti teorici fondamentali della antropologia di Kapp, si è passati ad analizzare le teorie e le prime architetture di computer intelligenti nei lavori di Turing e von Neumann. Si è scelto di non discutere in questa sede gli sviluppi successivi della AI, ciò sarebbe una impresa degna di una pubblicazione più vasta vista la complessità dell'argomento. Il riferimento a Kapp è servito per tratteggiare una teoria della AI che, invece di interpretare l'organismo *sub speciae machinae*, intende invertire il rapporto e interpretare le macchine come proiezioni di funzioni organiche. Tale rapporto di identificazione è ontologicamente secondo Kapp sempre difettivo e sbilanciato dal lato dell'organismo. Ciò permette di opporre al riduzionismo e al meccanicismo un paradigma interpretativo che considera l'organismo vivente secondo la sua

specificità e non in maniera difettiva. Così facendo, la questione della AI è contestualizzabile all'interno di una cornice più generale di rapporti tra organismo e tecnica, tra filosofia della tecnica e filosofia della biologia. Se la macchina nella storia della AI ha spesso costituito il modello interpretativo dell'intelligenza, qui si tratta proprio di invertire il rapporto, di considerare le macchine intelligenti come proiezioni ed estensioni della mente.

Questo lavoro sancisce un primo passo verso questa direzione. Ad ogni modo, sia il riduzionismo che l'organicismo della teoria della proiezione organica lasciano aperta la strada alla possibilità di attribuire alla macchina una qualche forma di intelligenza in grado di entrare in una relazione informazionale con l'uomo. Le obiezioni di Searle, di Edelman discussa in nota e di voci come quella di Hubert Dreyfus, che non si è avuto modo di discutere ampiamente per questione di spazio⁶³, per chi scrive, centrano il punto: non è possibile pensare un'intelligenza generale disincarnata di tipo umano e biologico al di fuori dell'essere radicato della vita nel mondo e al di là della sua connotazione sistemico-ecologica. A ciò aggiungiamo che qualsiasi linguaggio, *in primis* quello matematico e della logica formale sono atti creativi umani troppo umani che, pur costituendo delle straordinarie tecniche per decifrare e prevedere il mondo e creare strumenti tecnici sempre più sofisticati, sono sempre rappresentazioni, metonimie, strumenti e metafore di processi, più che descrizioni di fatti oggettivi. In poche parole, la riduzione del pensiero in calcolo o in processi logici formalizzabili secondo le regole della logica formale non possono rispecchiare il reale funzionamento del cervello biologico e della intenzionalità, ma ne costituiscono piuttosto una rappresentazione, un modello interpretativo valido in determinati ambiti. Da questo punto di vista, la prospettiva antiriduzionista e biologista che sostiene l'irriducibilità della macchina all'organismo appare ad oggi un forte argomento contro la realizzazione della AGI. Eppure, questo non ci permette di liquidare semplicemente la questione.

In primo luogo, la AI è a tutti gli effetti il risultato, più o meno esplicito, di una proiezione di funzioni organico-cognitive, come ad esempio il calcolo, la gestione di sistemi, il funzionamento neuronale, in una realtà artificiale; proiezione che, si noti bene, preparano un potenziamento ma anche una disattivazione (*Ausschaltung*) e un esonero dell'umano a svolgere determinate operazioni. In questo senso, le macchine relativamente intelligenti stanno sempre di più soppiantando l'umano nella gestione dei processi automatici o automatizzabili, robot capaci di svolgere mansioni umane sono oramai una realtà, così come computer quantistici in grado di interpolare immense moli di Big Data hanno per sempre cambiato la ricerca scientifica, implementando il nostro potere computazionale e predittivo. Insomma, noi viviamo in un ambiente tecnologico dove la simbiosi tra organico ed inorganico, naturale e artificiale, umano e macchinico è sempre più pervicace, al punto che oggi ci troviamo a condividere

⁶³ Per una trattazione della teoria di Dreyfus si rimanda a E. Carli, F. Grigenti, *Mente, cervello, intelligenza artificiale*, pp. 71-81, e all'opera stessa del filosofo: *Che cosa non possono fare i computer. I limiti della intelligenza artificiale*, tr. it. G. Alessandrini, Roma 1998.

il nostro spazio fisico e comunicativo con enti informazionali non umani. Il mondo di apparati contemporaneo è popolato da organismi informazionali che comunicano e interagiscono tra loro a prescindere dalla loro natura organica o artificiale. Come sosteneva Latour nella sua *Actor Network Theory* ogni relazione sociale vede l'interazione di fattori umani e non umani e oggi molti di questi non umani sono agenti intelligenti artificiali (*Intelligent Agent*)⁶⁴, che partecipano attivamente alla nostra galassia informazionale. Le intelligenze artificiali fanno oramai parte del nostro quotidiano.

Da questo punto di vista, sebbene si sia ancora lontani da una prospettiva sci-fi a là Blade Runner, dall'avvento di una singolarità e della AGI, la necessità di un paradigma critico-ermeneutico per la AI che orienti tanto il nostro agire etico che la ricerca teorica appare quantomai necessario.

La prospettiva di Kapp aiuta a contestualizzare i rapporti tra organismo e artificio a partire dalla centralità dell'organismo. Questo approccio ha il vantaggio di non schiacciare la questione della macchina a mero rispecchiamento e modello per l'organico, ma conserva per quest'ultimo una specificità e un'eccedenza che nessuna proiezione colma in maniera definitiva. In questo senso, esso ci aiuta nell'epoca della tecnica e della generale reificazione e (ri)produzione dell'ente e del mondo, in cui tutto anche l'organismo e l'uomo in quanto tale sono considerati come materia e fondo riproducibile e impiegabile nel sistema di apparati, a segnalare sempre uno scarto tra l'ente tecnico e l'organismo di cui è proiezione. Una conseguenza della teoria della proiezione come teoria genealogica e ontologia dell'ente tecnico ci mette al riparo anche dalla presunta antiquatezza dell'umano rispetto alla macchina, questione quanto mai attuale in un'epoca in cui le intelligenze artificiali sembrano in grado di effettuare qualunque compito in maniera più veloce, esaustiva e potente di qualsiasi intelligenza organica.

Turing e von Neumann hanno inaugurato una nuova frontiera nella nostra cultura materiale, che dopo alti e bassi, inverni ed entusiasmi ha per sempre modificato la nostra relazione con il mondo, con gli oggetti e con noi stessi. La AI realizza una nuova fase specifica della nostra storia evolutiva, l'epoca in cui alcuni aspetti e funzioni dell'intelligenza umana sono sempre più esteriorizzati e potenziati nel macchinico. Ciò equivale a proiettare nell'ente tecnico il nostro *principium individuationis*, come descritto nella parte introduttiva di questa trattazione. È naturale che di fronte alla epocalità di tale cambiamento l'uomo possa sentirsi detronizzato e privato di quella categoria, l'intelligenza, che ne aveva costituito il criterio specifico di autointerpretazione in tutta la storia della metafisica occidentale. Di fronte a tale cambiamento l'atteggiamento non può essere né eccessivamente apocalittico né acriticamente apologetico, in quanto

⁶⁴ Per *Intelligent Agent* si intende un qualsiasi agente che abbia una relazione con l'ambiente circostante attraverso dei sensori e agisca attraverso attuatori. Nel campo della AI un agente intelligente è una macchina che sappia interagire e adattarsi alle mutazioni del contesto, sappia eseguire compiti non necessariamente codificati e che sia in qualche modo capace di apprendere. Gli schemi di base di tali agenti sono comunque identificabili come derivazione della architettura di von Neumann.

la AI è latrice di opportunità e rischi. In primo luogo, essa ha costituito uno stimolo, un modello e uno strumento per le neuroscienze per dischiudere nuove prospettive sui processi cognitivi, essa ha rivoluzionato la ricerca scientifica, sta rivoluzionato il mondo del lavoro, della ricerca e anche della giurisprudenza. Ha accresciuto il nostro potere predittivo, incrementando la nostra capacità di reagire a situazioni di crisi e pandemie, ma ha anche inverato la possibilità di un nuovo totalitarismo della sorveglianza⁶⁵ e la perdita della privacy. Inoltre, è sempre più frequente la delega della nostra facoltà decisionale al potere degli algoritmi⁶⁶. La AI ha poi posto nuove questioni etiche e giuridiche circa il suo uso e regolamentazione⁶⁷.

Tuttavia, a ben vedere, il fatto che le macchine riescano a imparare, a guidare automobili autonomamente, battere campioni di scacchi, a passare il test di Turing e imitare funzioni cognitive umane come nel caso di LaMDA e ChatGPT, ci dice molto di più quanto sia straordinaria l'intelligenza umanissima degli ingegneri informatici, che l'effettiva intelligenza della macchina, almeno finché una macchina non sarà in grado di generare un'altra macchina intelligente. Il che equivale a sottolineare quanto sia singolare il modo di abitare tecnicamente il mondo di quell'animale, nato senza un luogo e una dotazione naturale specifica, che incessantemente proietta se stesso e le sue funzioni nella materia bruta del mondo facendone il terreno della sua exteriorizzazione. Sarebbe pertanto utile approcciare la questione della intelligenza artificiale, conformemente allo schema della proiezione organica, come un modo dell'agire umano in senso globale e non esclusivamente dell'intelligenza⁶⁸, mantenendo sempre ben presente l'eccedenza e l'eccentricità che ogni organismo in via di principio conserva nei confronti delle sue proiezioni. Solo in questo modo la centralità dell'umano può essere ribadita rispetto allo strapotere del mondo degli apparati (*Apparatenwelt*). In conclusione, le conseguenze epistemologiche di questo approccio sono ormai chiare: le macchine, tra cui la AI, sono un eccezionale strumento ermeneutico per arrivare all'autocoscienza della nostra dotazione organica e per implementare la nostra conoscenza e presa sul mondo; inoltre, il vivente costituisce quel termine *ad quem* cui l'artificio tenderebbe, ancora nel caso dell'AGI come riproduzione delle facoltà cognitive umane. La macchina non genera macchine, e anche se in futuro dovesse farlo, questa capacità non sarebbe altro che il rispecchiamento di una facoltà umana. Pertanto, per quanto sofisticato, potente e pervasivo potrà

⁶⁵ Cfr. S. Zuboff, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, tr. it. P. Bassotti, Roma 2019.

⁶⁶ Cfr. M. Durante, *Potere computazionale, L'impatto delle ICT su dritta società e sapere*, Milano 2019 e C. O'Neil, *Armi di distruzione matematica. Come i big data aumentano la disuguaglianza e minacciano la democrazia*, tr. it. D. Cavallini, Bompiani, Milano 2017.

⁶⁷ Ne è testimonianza la *AI Act* della Commissione Europea volta a regolamentare l'uso della AI: <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>.

⁶⁸ A tale conclusione arriva Luciano Floridi anche se a partire da un'impostazione teorica totalmente differente estranea alla antropologia filosofica nel suo *Etica della intelligenza artificiale. Sviluppi opportunità sfide*, Milano 2022.

divenire il nostro mondo di apparati sarà sempre un rispecchiamento, più o meno inconscio, della nostra fisiologia e quindi sua *eikona*, *mimesis* e copia, da cui deve necessariamente trarre la misura.

Lorenzo De Stefano
Università degli Studi di Napoli Federico II
✉ lorenzo.destefano@unina.it

Bibliografia

- Alsberg, P. 2021. *L'enigma dell'umano. Per una soluzione biologica*, tr. it. E. Nardelli, Roma, Inschibboleth.
- Anders G. 2007. *L'uomo è Antiquato, vol. 1: Considerazioni sull'anima nell'epoca della seconda rivoluzione industriale*, tr. it L. Dallapiccola, Torino, Bollati Boringhieri.
- Aristotele 1990. *Parti degli animali*, a cura di M. Vegetti, in *Opere*, vol. 5, Roma-Bari, Laterza.
- Ashby, W. R. 1947. *The Nervous System as Physical Machine: With Special Reference to the Origin of Adaptive Behavior*, «Mind, New Series» 56, 221, pp. 44-59.
- Bhattacharya, A., 2022. *The Man from the Future*, London, Penguin Books.
- Boden, M. A. 2019. *L'Intelligenza Artificiale*, a cura di D. Marconi, Bologna, Il Mulino.
- Bostrom N. 2018, *Superintelligenza. Tendenze, pericoli, strategie*, tr. it. S. Frediani, Torino, Bollati Boringhieri.
- Brown, F. 1992. *Tutti i racconti (1950-1972)*, tr. it. Aa.Vv., Milano, Mondadori.
- Carli, E., Grigenti, F. 2019. *Mente, cervello, intelligenza artificiale*, Milano-Torino, Pearson Italia.
- Coeckelbergh, M. 2022. *AI Ethics*, Princeton, MIT Press.
- Davis, M. 2012. *Il calcolatore universale. Da Leibniz a Turing*, tr. it G. Rigamonti e A. La Rana, Milano, Adelphi.
- Descartes, R. 2009. *Opere 1637-1649*, a cura di G. Belgioioso, Milano, Bompiani.
- Dupuy, J. P. 2015. *Alle origini delle scienze cognitive. La meccanizzazione della mente*, a cura di P. Heritier, Milano-Udine, Mimesis.
- Dreyfus H. 1998. *Che cosa non possono fare i computer. I limiti della intelligenza artificiale*, tr. it. G. Alessandrini, Roma, Armando Editore.
- Durante, M. 2019. *Potere computazionale, L'impatto delle ICT su dritta società e sapere*, Milano, Meltemi.
- Edelman, G. M. 2018. *Darwinismo neurale. La teoria della selezione dei gruppi neuronali*, a cura di S. Ferraresi, Raffaello Cortina Editore, Milano.

- Esposito, E. 2022. *Comunicazione artificiale. Come gli algoritmi producono intelligenza sociale*, Milano, Bocconi University Press.
- Floridi, L. 2017. *La Quarta Rivoluzione*, Milano, Raffaello Cortina Editore.
- Id. 2022. *Etica della intelligenza artificiale. Sviluppi opportunità sfide*, Raffaello Cortina Editore, Milano.
- Fodor, J. A. 2001. *Mente e linguaggio*, a cura di F. Ferretti, Bari, Laterza.
- Gigerenzer, G. 2023. *Perché l'intelligenza umana batte ancora gli algoritmi*, tr. it. R. Mazzeo, Milano, Raffaello Cortina Editore.
- Gilson, E. 1973. *La filosofia del medioevo. Dalle origini patristiche alla fine del XIV secolo*, Firenze, La Nuova Italia.
- Garber, D. 2002. *Descartes, Mechanics, and the Mechanical Philosophy*, «Midwest Studies in Philosophy», 26, Minnesota, Philosophy Documentation Center, pp. 185-204.
- Godfrey-Smith, P. 2012. *Metazoa. Gli animali e la nascita della mente*, tr. it. I. C. Blum, Milano, Adelphi.
- Guidi, S. 2018. *L'angelo e la macchina. Sulla genesi della res cogitans cartesiana*, Roma, Franco Angeli.
- Gottfredson, L. S. 1997. *Mainstream Science on Intelligence: An editorial with 52 signatories, history and bibliography*, «Intelligence» volume 24, 1, Delaware, Elsevier, pp. 13-23.
- Grigenti, F. 2021. *Le macchine e il pensiero*, Napoli, Orthotes.
- Heidegger, M. 1976. *Saggi e Discorsi*, a cura di G. Vattimo, Milano, Mursia.
- Haugeland J. 1985. *Artificial Intelligence: The Very Idea*, Princeton, MIT Press.
- Heims, S. J. 1994. *I cibernetici. Un gruppo e un'idea*, Roma, Editori Riuniti.
- Kapp, E. 2015. *Grundlinien einer Philosophie der Technik*, Hamburg, Felix Meiner Verlag.
- Kurzweil, R. 2018 *La singolarità è vicina*, Milano, Apogeo Editore.
- Malafouris, L. 2013. *How Things Shape the Mind: A Theory of Material Engagement*, Cambridge, MIT Press.
- McCulloch, W. & Pitts, W 1943. *A logical calculus of the Ideas Immanent in Nervous Activity*, «Bulletin of Mathematical Biophysics», 5, Oxford, Pergamon Press, pp. 115-133.
- M. McLuhann, M. 2001. *Understanding media. The extension of man*, London, Routledge.
- O'Neil, C. 2017 *Armi di distruzione matematica. Come i big data aumentano la disuguaglianza e minacciano la democrazia*, tr. it. D. Cavallini, Milano, Bompiani.
- Rumelhart, D. E., McClelland J. L. & the PDP Research Group 1986. *Parallel Distributed Processing: Explorations*, «Microstructure of Cognition» Volume 1: Foundations, Cambridge, 1, MIT Press.
- Searle, J. R. 1980. *Minds, Brains and Programs*, «Behavioral and Brain Sciences», 3, Cambridge, Cambridge University Press, pp. 417-424.
- Id. 2017. *La riscoperta della mente*, Torino, Bollati Boringhieri.

- Turing, A. 1936. *On computable Numbers, with an Application to the Entscheidungsproblem*, «Proceedings of the London Mathematical Society», London, LMS Publications, pp. 230-265.
- Id. 1950. *Computer Machinery and Intelligence*, «Mind», 56, Oxford, Oxford University Press, pp. 433-460.
- Von Neumann, J. 2021. *Computer e cervello*, tr. it. P. Bartesaghi, Milano, il Saggiatore.
- Wiener, N. 2012 *Introduzione alla cibernetica. L'uso umano degli esseri umani*, tr. it. D. Persiani, Torino, Bollati Boringhieri.
- Zuboff, S. 2019. *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, tr. it. P. Bassotti, Roma, LUISS University Press.