



UNIVERSITÀ DEGLI STUDI DI PADOVA

**DIPARTIMENTO DI FILOSOFIA, SOCIOLOGIA, PEDAGOGIA E
PSICOLOGIA APPLICATA**

CORSO DI LAUREA IN FILOSOFIA

”Il libero arbitrio nell’era dell’intelligenza artificiale: da Cartesio a Alan Turing”

Relatore:
[Ch.mo](#) Prof. Fabio Grigenti

Laureanda:
Letizia Andretta
Matricola n.2079122

ANNO ACCADEMICO 2025-2026

Introduzione

Capitolo 1 - Il libero arbitrio in Cartesio

1.1 Il progetto cartesiano

1.2 Intelletto e volontà: le due facoltà dell'anima e la libertà nella Quarta Meditazione

Capitolo 2 - Dalla macchina Cartesiana all'intelligenza artificiale: l'autonomia del pensiero nell'era tecnologica

2.1 Dal soggetto pensante alla macchina pensante

2.2 La nascita dell'idea di macchina intelligente

2.3 L'intelligenza artificiale: definizioni e principi generali

2.4 La continuità e la rottura con la visione cartesiana

2.4.1 Continuità con la visione

2.4.2 Rottura con la visione

2.4.3 Implicazioni per il libero arbitrio

Capitolo 3 - Il libero arbitrio nell'era dell'intelligenza artificiale

3.1 Cartesio e l'intelligenza artificiale

3.2 Introduzione: la questione della libertà nell'età digitale

3.3 La decisione tra volontà e calcolo

3.3.1 Intelligenza artificiale e notizie false

3.4 Autonomia e determinismo nelle macchine intelligenti

3.5 Libertà come emergenza o simulazione?

3.6 Conclusione: l'uomo di fronte alla macchina: chi è ancora libero?

Capitolo 4 - Etica, responsabilità e pensiero contemporaneo nell'era dell'intelligenza artificiale e in Cartesio

4.1 Introduzione: la questione morale dopo l'automazione del pensiero

4.2 Libertà e responsabilità: l'uomo davanti alle decisioni delle macchine

4.3 Prospettive etiche contemporanee

4.3.1 Etica della progettazione delle macchine intelligenti: criteri,
limiti e responsabilità degli sviluppatori

4.3.2 Intelligenza artificiale e legislazione

4.4 Etica e responsabilità in Cartesio

Capitolo 5 - Conclusioni

Introduzione

La questione del libero arbitrio costituisce uno dei nodi più antichi e complessi della riflessione filosofica. La libertà dell'uomo, intesa come capacità di scegliere consapevolmente tra alternative possibili e di agire secondo ragione, ha da sempre implicazioni morali, etiche e metafisiche. Fin dall'antichità, filosofi come Agostino, Tommaso d'Aquino e Aristotele hanno discusso i limiti e le condizioni della libertà umana, cercando di conciliare la responsabilità individuale con la struttura dell'universo e con eventuali vincoli divini o naturali. Nel corso della modernità, la riflessione sul libero arbitrio subisce una trasformazione radicale con René Descartes, che propone una concezione dell'uomo come essere fondamentalmente pensante, capace di autodeterminarsi attraverso l'esercizio della ragione. La volontà, in Cartesio, non è semplicemente un impulso o un desiderio, ma una facoltà che permette all'uomo di orientare le proprie scelte verso la verità e il bene, pur operando all'interno di un mondo ordinato e meccanico.

La presente tesi si propone di analizzare il concetto di libero arbitrio secondo Cartesio e di esplorare le implicazioni di tale concetto nell'era contemporanea, caratterizzata dalla rapida diffusione delle macchine intelligenti e dei sistemi di intelligenza artificiale. La riflessione cartesiana sulla volontà e sulla coscienza umana offre un punto di partenza privilegiato per confrontare la libertà dell'uomo con le capacità decisionali delle macchine. La domanda di fondo che guida il lavoro è: cosa significa essere liberi oggi, in un contesto in cui la razionalità può essere simulata, replicata e talvolta superata dalle macchine?

Il percorso della tesi si sviluppa in cinque capitoli, ciascuno con un obiettivo preciso e collegato agli altri. Il **Capitolo 1** si concentra sul libero arbitrio in Cartesio, analizzando il progetto cartesiano, la distinzione tra intelletto e volontà, e la concezione della libertà come autodeterminazione razionale delineata nella Quarta Meditazione. Questo capitolo non ha soltanto la funzione di spiegare le idee cartesiane, ma anche di porre le basi concettuali per il confronto con le macchine intelligenti, evidenziando i fondamenti filosofici del libero arbitrio e i limiti imposti dalla natura umana e dal contesto metafisico cartesiano.

Il **Capitolo 2** apre la prospettiva contemporanea, collegando la visione cartesiana dell'uomo come soggetto pensante alla nascita e allo sviluppo dell'intelligenza artificiale. Si analizzeranno le origini dell'idea di macchina intelligente, a partire dai primi dispositivi meccanici capaci di imitare azioni umane, del Seicento, fino alla disciplina dell'IA moderna, delineando definizioni e principi generali che ne regolano

il funzionamento. In questo contesto, particolare attenzione sarà riservata ad Alan Turing, il quale ha formalizzato il concetto di macchina capace di calcolare e ragionare secondo regole logiche, introducendo strumenti teorici fondamentali come la macchina di Turing e il celebre test di Turing per valutare la capacità di un sistema artificiale di simulare l'intelligenza umana. Verrà inoltre discusso il rapporto tra continuità e rottura rispetto alla visione cartesiana, mostrando come la macchina possa riprodurre alcune funzioni razionali, ma senza possedere autonomia cosciente o volontà morale. Questo capitolo ha la funzione di introdurre strumenti e concetti indispensabili per valutare la libertà nell'era digitale.

Nel **Capitolo 3** si entra nel cuore del problema contemporaneo: la libertà nell'era dell'intelligenza artificiale. Si analizzerà la differenza tra la decisione umana, fondata sulla volontà e sulla coscienza, e le decisioni algoritmiche, basate su regole, dati e probabilità. Si discuteranno concetti come autonomia e determinismo nelle macchine intelligenti, oltre a interrogarsi sulla possibilità che emergano forme di libertà o simulazioni sofisticate di essa. Infine, il capitolo proporrà una riflessione sul ruolo dell'uomo di fronte alla macchina, interrogandosi su chi sia realmente libero quando la tecnologia può influenzare profondamente le scelte quotidiane.

Il **Capitolo 4** affronta le implicazioni morali e etiche dell'era digitale, considerando i contributi di filosofi contemporanei che riflettono su responsabilità e libertà in un mondo dominato dalle macchine. Si prenderanno in esame autori come Luciano Floridi, Daniel Dennett e John Searle, che discutono della libertà come responsabilità relazionale, della simulazione della coscienza nelle macchine e della possibile erosione della libertà umana nell'infosfera digitale. Questo capitolo permette di integrare l'analisi filosofica con riflessioni etiche, mostrando come il tema del libero arbitrio non sia più confinato alla metafisica, ma investa decisioni concrete, normative e sociali.

Infine, il **Capitolo 5** costituisce la chiusura della tesi, con le conclusioni e le riflessioni personali. Qui si tratteranno i risultati principali, evidenziando come la libertà, pur trovando nuove sfide nell'era dell'intelligenza artificiale, rimanga un elemento centrale della condizione umana. Si discuterà di come la libertà non sia semplicemente assenza di vincoli, ma capacità di orientarsi razionalmente e moralmente, preservando l'autonomia in un contesto tecnologico complesso. Le considerazioni finali offriranno una prospettiva critica e personale sul futuro della libertà e della responsabilità nell'interazione tra uomo e macchina.

La tesi, nel suo insieme, intende mostrare come il concetto di libero arbitrio non sia un tema astratto o confinato alla filosofia classica, ma un problema attuale e urgente, la cui analisi richiede l'integrazione tra storia del pensiero, filosofia della mente, scienze cognitive e riflessione etica contemporanea. Il filo conduttore che lega Cartesio alle macchine intelligenti è la domanda fondamentale sulla natura della libertà: cosa

significa essere liberi, e in che modo possiamo comprendere e tutelare questa libertà quando i processi decisionali non appartengono più soltanto all'uomo, ma si estendono a sistemi artificiali sempre più complessi?

In questo senso, la tesi si pone come un ponte tra passato e presente, tra metafisica e tecnologia, tra filosofia classica e sfide contemporanee, cercando di offrire una prospettiva critica e coerente sul tema centrale del libero arbitrio. Il lettore sarà così guidato lungo un percorso che parte dai fondamenti cartesiani della libertà e giunge alle questioni più attuali legate all'intelligenza artificiale, passando per riflessioni etiche, morali e sociali che rendono il tema della libertà sempre più urgente e rilevante.

Capitolo 1 - Il libero arbitrio in Cartesio

1.1 Il progetto Cartesiano

René Descartes (1596-1650) è comunemente riconosciuto come il fondatore della filosofia moderna, a certificarlo sono da sempre i filosofi stessi, in quanto troviamo la sua impronta in ogni tematica cruciale dell'epoca moderna¹ e poiché inaugura un nuovo modo di intendere la conoscenza e il rapporto dell'uomo con la verità.

Il suo progetto filosofico nasce dall'esigenza di dare alla scienza un fondamento certo, capace di superare ogni incertezza della tradizione scolastica e della conoscenza sensibile.

Nel Discorso sul metodo (1637), Descartes formula il principio cardine del suo pensiero, ovvero quello che la ragione è l'unico strumento in grado di condurre alla verità, purché segua un metodo rigoroso. Il dubbio è quindi il punto di partenza di ogni certezza e conoscenza: bisogna mettere in questione ogni opinione, fino a trovare qualcosa che resista ad ogni tentativo di confutazione. Tale certezza si manifesta nella celebre intuizione del: « Cogito ergo sum, sive existo »² - « Io penso, dunque sono, ossia esisto »-, che diventa il fondamento di ogni verità possibile.

L'intero progetto Cartesiano ha come obiettivo il rifondare il sapere sulla base dell'autonomia della ragione in cui l'uomo attraverso il solo pensiero, non dipende più dall'autorità esterna o dalla tradizione, ma diventa il centro del conoscere.

Descartes scrive infatti: « Tutta la filosofia è come un albero, le cui radici sono la metafisica, il tronco è la fisica e i rami tutte le altre scienze »³, immagine che esprime l'unità e la gerarchia del sapere fondato sul metodo.

All'interno di questo progetto, la libertà dell'uomo assume un ruolo significativo, infatti, per il filosofo l'uomo possiede due facoltà: l'intelletto, che conosce ed è limitato e la volontà che sceglie ed è praticamente illimitata. La libertà invece è proprio ciò che consente all'uomo di autodeterminarsi e di orientarsi verso la verità; tuttavia, la sua piena realizzazione si ottiene solo quando la volontà segue l'intelletto e aderisce a ciò che appare chiaro e distinto. In tal modo, la libertà non è semplice indifferenza, ma adesione consapevole alla ragione, essa rappresenta il punto d'incontro tra conoscenza e moralità e come vedremo in modo più approfondito nei capitoli successivi, proprio questa concezione di libertà come esercizio razionale, costruirà la base per il confronto con le forme di "autonomia artificiale" dell'età contemporanea.

¹ Gianluca Mori, *Cartesio*, Carocci editore, 2010, p.9.

² Descartes, R., *Discourse de la Méthode* (1637), Mondadori, Milano 1993, pp. 34.

³ Descartes, R., Prefazione ai principi di filosofia.

1.2 Intelletto e volontà: le due facoltà dell'anima

Nella Quarta Meditazione, Cartesio affronta direttamente la questione della verità e dell'errore umano, collegandola al concetto di libertà.

Particolarmente rilevante è la distinzione tra intelletto e volontà, che ci aiuta a comprendere la natura dell'anima e il funzionamento della libertà umana. L'anima è definita come *res cogitans*, ed è la sostanza pensante la cui essenza consiste nel pensare in tutte le forme: comprendere, volere, immaginare e sentire. Tuttavia, tra queste attività se ne distinguono due principali: l'intelletto e la volontà.⁴

L'intelletto è per Cartesio, una facoltà limitata, infatti esso può percepire le idee ma non ne produce liberamente il contenuto, più precisamente esso è lo strumento mediante cui l'uomo accede alla conoscenza, ma la sua portata rimane finita⁵. L'intelletto si limita a presentare alla mente ciò che è oggetto di pensiero, senza ancora esercitare un giudizio.

La volontà, invece, viene descritta come illimitata, infatti in particolare nella quarta meditazione viene evidenziata la sproporzione tra le due facoltà: l'intelletto umano è finito e circoscritto, mentre la volontà si estende senza confini, potendo affermare o negare qualsiasi cosa, anche ciò che non è compreso pienamente⁶.

Cartesio individua proprio in questa sproporzione la radice dell'errore umano, esso nasce quando la volontà si spinge a giudicare questioni che l'intelletto non percepisce con chiarezza e distinzione⁷. Dunque, l'intelletto non può sbagliare, poiché percepisce soltanto, la volontà, invece, può andare oltre ai limiti della conoscenza scegliendo di assentire a idee non chiare e distinte. In altre parole, l'errore non è un difetto naturale delle facoltà, ma il risultato di un uso incorretto della libertà.

La vera libertà non consiste dunque nel poter scegliere indifferentemente tra più alternative, ma nella capacità di aderire in modo consapevole al vero, quando la volontà segue ciò che l'intelletto riconosce come chiaro e distinto, l'uomo agisce, oltre che in modo libero, anche in modo corretto. Come scrive Cartesio ad Elisabetta di

⁴ Descartes, R., *Principia Philosophiae*, vol. VIII, p. 17.

⁵ «L'intelletto non è altro che la facoltà di conoscere, che non afferma né nega nulla», Descartes R., *Principia Philosophia*.

⁶ D. Garber, *Descartes' Metaphysical Physics*, Chicago, University of Chicago Press, 1992, p.134.

⁷ «L'errore per Cartesio dipende da due cause tra loro convergenti, dalla *facultas cognoscendi* e dalla *facultas eligendi*, cioè dall'intelletto e dalla volontà; più precisamente dal giudizio che asseconda l'esigenza della volontà di trascendere i limiti dell'intelletto», Introduzione a Renato Cartesio, *Meditazioni metafisiche*, Armando Editore, 2003, p.26.

Boemia⁸: « In questa sola libertà positiva, cioè nella facoltà di seguire il vero, risiede la perfezione della volontà »⁹.

L'uomo è dunque realmente libero perché può scegliere in modo arbitrario, unendo il sapere al volere e nella misura in cui unisce la chiarezza dell'intelletto alla rettitudine della volontà¹⁰.

La distinzione tra intelletto e volontà nel pensiero cartesiano non segna una distinzione, ma una tensione dinamica che definisce la condizione umana, l'uomo è libero nella misura in cui unisce la chiarezza dell'intelletto alla rettitudine della volontà. Tale equilibrio, fondamento della libertà morale, costituirà nei capitoli successivi, il punto d'incontro di confronto con le nuove forme di decisione automatica e pilotata, proprie dell'intelligenza artificiale.

⁸ Elisabetta di Boemia (1618-1680) è una figura centrale nella vita e nel pensiero di René Descartes, non solo per il loro intenso scambio epistolare, ma anche per il ruolo che ella ebbe nel mettere alla prova e approfondire alcuni punti cruciali della sua filosofia. Ella era la figlia dell'elettore palatino Federico V e di Elisabetta Stuart d'Inghilterra, appartenente dunque a una delle più illustri famiglie europee del tempo. Le lettere tra Cartesio ed Elisabetta affrontano temi fondamentali della metafisica e dell'etica cartesiana: il problema dell'unione tra anima e corpo, la natura delle passioni, il rapporto tra ragione e volontà, e infine il senso della libertà.

⁹ Descartes, R., *Epistolae ad Elisabet*, 6 ottobre 1645, IV, 266.

¹⁰ Descartes R., *Meditationes*, IV.

Capitolo 2- Dalla macchina Cartesiana, all'intelligenza artificiale: l'autonomia del pensiero nell'era tecnologica

2.1 Dal soggetto pensante alla macchina pensante

L'eredità del pensiero cartesiano, basato sull'ideale di una conoscenza certa fondata sul metodo, ha condizionato profondamente lo sviluppo del pensiero scientifico e tecnologico moderno e ha dato origine ad una visione dell'uomo come soggetto pensante contrapposto a un mondo esterno concepito come meccanismo regolato da leggi deterministiche.

La distinzione cartesiana tra *res cogitans* e *res extensa* ha fondato il paradigma meccanicistico moderno, in cui la mente è pensiero e la materia puro movimento, rendendo così il mondo fisico analizzabile e replicabile secondo modelli razionali e matematici¹¹.

Se la mente è pensiero e la materia puro movimento, allora il corpo, e il mondo fisico, possono essere studiati e replicati secondo modelli razionali e matematici¹². Questa riduzione del vivente a macchina, già anticipata da Cartesio nel *Traité de l'homme*, dove descrive il corpo umano come un automa animato dal movimento degli spiriti vitali¹³, apre la strada alla possibilità di pensare la macchina non solo come imitazione della natura, ma come suo possibile prolungamento.

Con l'età moderna, il progetto cartesiano viene progressivamente secolarizzato e tradotto in termini tecnico-scientifici, specialmente nel Settecento e nell'Ottocento, il modello meccanico della natura ispira la nascita delle prime teorie sull'automatismo e sull'automa come dimostrano le celebri costruzioni di Vaucanson e di Kempelen¹⁴. La macchina diventa non solo strumento di lavoro ma anche simbolo dell'intelligenza umana applicata alla materia, un riflesso del cogito stesso.

L'uomo, creatore di meccanismi, riproduce nella tecnica la propria razionalità: la macchina diventa, in un certo senso, "pensante", perché agisce secondo regole determinate e reperibili.

Con l'avvento del XX secolo, questa linea di continuità tra il meccanismo cartesiano e le macchine moderne si realizza nella cibernetica e nelle teorie dell'informazione.

¹¹ R.Descartes, *Meditazioni metafisiche*, trad. Di E. Grain, Laterza, 1997, p. 77.

¹² Stephen Gaukroger, *Descartes. An intellectual Biography*, Oxford, Clarendon Press, 1995, p. 221.

¹³ R. Descartes, *Trattato dell'uomo*, pp. 120-125.

¹⁴ Jessica Riskin, *The Restless Clock: A History of the Centuries-Long Argument over What Makes Living Things Tick*,

University of Chicago Press, 2016, pp. 33-37. Gli automi descritti da Riskin, come l'anatra digerente di Vaucanson o il Turco di Kempelen, erano macchine meccaniche progettate per imitare funzioni biologiche o comportamenti umani e animali. Non possedevano coscienza né intelligenza reale, ma eseguivano movimenti complessi secondo schemi meccanici prestabiliti, dimostrando la possibilità di replicare fenomeni naturali tramite ingranaggi e leve.

Norbert Wiener, nel 1948, definisce la cibernetica come «la scienza del controllo e della comunicazione nell'animale e nella macchina. »¹⁵, ponendo per la prima volta un piano di equivalenza funzionale tra organismo e automa. La mente umana e la macchina elettronica vengono così considerate due sistemi capaci di elaborare informazioni, seppur in forme differenti. L'intelligenza non è più un privilegio dell'anima, ma un processo che può essere descritto, misurato e forse replicato¹⁶.

In questa prospettiva, l'uomo non è più soltanto soggetto pensante, ma anche progettista di una nuova forma di pensiero: quella artificiale. La transizione dalla *res cogitans* alla "macchina pensante" non rappresenta soltanto una trasformazione tecnica, ma una svolta ontologica: ciò che era un tempo prerogativa della mente umana, ovvero il pensare, diventa un attributo condiviso con l'artificiale¹⁷.

2.2 La nascita dell'idea di macchina intelligente

L'idea di macchina intelligente emerge con la crescente comprensione con la crescente comprensione dei processi logici e matematici alla fine del XIX e all'inizio del XX secolo. In questo periodo, diversi matematici e ingegneri iniziarono a concepire dispositivi capaci di eseguire operazioni complesse senza l'intervento umano diretto, aprendo la strada alla distinzione tra automi meccanici e sistemi pensanti¹⁸.

Un ruolo di rilievo è svolto da George Boole, il cui lavoro sull'algebra permise di tradurre il ragionamento umano in simboli e operazioni logiche, aprendo la strada alla formalizzazione dei processi decisionali e alla possibilità di costruire macchine capaci di manipolare simboli secondo determinate regole¹⁹.

Successivamente, nel XX secolo, Alan Turing ampliò radicalmente questa prospettiva introducendo il concetto di macchina universale, in grado di simulare qualsiasi calcolo computabile²⁰. Turing dimostrò che i processi logici e matematici potevano essere automatizzati, delineando un modello teorico che oggi chiamiamo intelligenza artificiale; tuttavia, egli sottolinea anche i limiti intrinseci: una macchina, per quanto complessa, segue regole definite e non possiede "volontà" autonoma, evidenziando una distinzione fondamentale rispetto alla libertà cartesiana dell'uomo²¹.

¹⁵ Norbert Wiener, *Cybernetics: or Control and Communication in the Animal and the Machine*, Cambridge, 1948, p.19.

¹⁶ Turing, A. M., "Computing Machinery and Intelligence", in *Mind*, vol. 59, n.236, 1950, p.433.

¹⁷ Luciano Floridi, *The Philosophy of Information*, Oxford University Press, 2011, p.67.

¹⁸ Wiener, N., *Cybernetics: Or Control and Communication in the Animal and the Machin*, MIT Press, 1948, pp. 15-22.

¹⁹ Boole, G., *An Investigation of the Laws of Thought*, Dover, 1958 [ed. Originale 1854], pp. 5-10.

²⁰ Turing, A. M., "On Computable Numbers, with an Application to the Entscheidungsproblem".

Proceedings of the London Mathematical Society, Series 2, vol 42. 1936, pp. 230-265.

²¹ Turing, A. M., op. Cit., pp. 247-250.

Parallelamente il concetto di macchina intelligente viene alimentato anche dalle intuizioni di Norbert Wiener e dalla cibernetica, che negli anni '40 e '50 introdussero l'idea di sistemi capaci di autoregolazione e apprendimento attraverso feedback²². Questa evoluzione concettuale segnò un passaggio cruciale: non si trattava più solo di calcolare o eseguire istruzioni, ma di creare sistemi in grado di adattarsi a situazioni nuove, avvicinandosi a ciò che oggi definiamo autonomia operativa delle macchine.

2.3 L'intelligenza artificiale: definizioni e principi generali

L'intelligenza artificiale, o IA, può essere definita come l'insieme di tecniche e sistemi informatici progettati per svolgere compiti che normalmente richiederebbero l'intelligenza umana. Questi compiti includono: l'elaborazione di informazioni, il riconoscimento di schemi, la risoluzione di problemi e, in alcuni casi, la capacità di apprendere dall'esperienza²³. In altre parole, l'IA non "pensa" come un uomo, ma simula alcune funzioni cognitive attraverso algoritmi e modelli matematici.

Un principio fondamentale dell'IA è la rappresentazione della conoscenza, ovvero la capacità di trasformare informazioni sull'ambiente in una forma che il sistema possa interpretare e usare per prendere decisioni²⁴. Questa rappresentazione può avvenire attraverso regole logiche, reti neurali o modelli statistici, ciascuno con vantaggi e limiti specifici.

Un altro principio importante è l'apprendimento automatico (machine learning), ovvero la capacità dei sistemi di implementare le proprie prestazioni grazie all'esperienza o alle informazioni a disposizione²⁵. Attraverso algoritmi supervisionati, non supervisionati o di rinforzo, le macchine possono individuare schemi e fare previsioni senza essere programmate per ogni compito specifico.

Centrale per l'IA è anche la capacità decisionale, grazie ad essa può selezionare azioni adeguate in base a obiettivi e vincoli noti. In alcuni sistemi, questa capacità si limita a semplici algoritmi di ricerca e organizzazione, in altri può includere modelli più complessi di ragionamento probabilistico o simulazioni predittive²⁶.

L'insieme di questi principi, sono il fondamento teorico e pratico dell'intelligenza artificiale moderna, essi forniscono strumenti per comprendere non solo come le

²² Wiener, N., op. Cit., pp. 45-60.

²³ Russel, S., & Norving, P. *Artificial Intelligence: A Modern Approach*, IV ed., Pearson, 2021, p.2.

²⁴ Luger, G. F., *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, Pearson, 2020.

²⁵ Goodfellow, I., Bengio Y., & Courville, A. *Deep Learning*, 2016, pp. 9-12.

²⁶ Poole, D., Mackworth, A., & Goebel, R. *Computational Intelligence: A Logical Approach*, Oxford University Press, 2017, pp.21-24.

macchine “pensano”, ma anche i limiti e le potenzialità dei sistemi che saranno al centro dei dibattiti successivi sul libero arbitrio e sulla responsabilità umana.

2.4 La continuità e la rottura con la visione cartesiana

In questo paragrafo si intende confrontare la concezione del soggetto e del libero arbitrio in René Descartes con le trasformazioni che emergono con l'avvento dell'intelligenza artificiale, individuando ciò che permane e ciò che invece viene superato.

2.4.1 Continuità con la visione

Cartesio, come evidenziato nei capitoli precedenti in particolare nella sezione 1.2, pone al centro della sua riflessione il soggetto pensante, fondamento della conoscenza e della libertà dell'anima²⁷, e le due facoltà distinte ovvero: l'intelletto che conosce e la volontà che sceglie²⁸.

In questa prospettiva è possibile rintracciare una continuità nella riflessione contemporanea sull'intelligenza artificiale, in quanto permane la nozione che la libertà, ossia la capacità di essere autore della propria azione, resti un tema centrale e che la soggettività non possa essere ridotta a mero calcolo o automazione.

Un secondo aspetto riguarda appunto la distinzione tra intelletto e volontà, questa distinzione cartesiana, pur essendo stata formulata nel XVII secolo, fornisce ancora oggi un paradigma utile per comprendere la relazione tra uomo e macchina: gli algoritmi possono essere visti come “intelletto” artificiale, capaci di elaborare informazioni e produrre risultati, mentre la volontà umana rimane il parametro critico attraverso cui decidere se accettare, modificare o rifiutare le decisioni prodotte dalla macchina. In tal senso, la riflessione cartesiana consente di leggere la tecnologia come estensione delle capacità cognitive dell'uomo senza che questo annulli la sua autonomia decisionale.

Un'ulteriore continuità riguarda la centralità della responsabilità, Cartesio infatti individua nella volontà umana la fonte della libertà ma anche dell'errore: infatti sbagliamo quando la volontà si estende oltre ciò che l'intelletto comprende²⁹. Analogamente, nell'era dell'intelligenza artificiale la responsabilità ultima delle decisioni non è attribuibile alle macchine stesse, ma al soggetto umano che le utilizza,

²⁷ R. Descartes, *Meditations on First Philosophy*, in J. Cottingham, Cambridge University Press, 1986, Meditazione II.

²⁸ Ivi, Meditazione IV, pp. 34-38.

²⁹ Ivi, pp. 35-37.

calibra, orienta e ne verifica i risultati. La nozione cartesiana di soggetto agente e responsabile conserva la sua rilevanza come principio guida nella progettazione e nell'uso dei sistemi intelligenti.

2.4.2 Rottura con la visione

La rottura con la visione cartesiana emerge con la trasformazione del ruolo dell'uomo nell'era digitale. Cartesio, infatti, poneva il soggetto come centro dell'agire ma le tecnologie intelligenti introducono una relazione molto complessa tra soggetto e sistemi autonomi³⁰. In questo contesto, la decisione non è più esclusivamente attribuibile all'uomo, ma diventa il risultato dell'interazione con algoritmi capaci di analizzare dati, proporre azioni o addirittura anticipare comportamenti. Ciò mette in discussione l'idea cartesiana di centralità e soprattutto unicità del soggetto agente.

Un'altra linea di rottura riguarda la possibilità di modellare e prevedere i processi decisionali. In Cartesio, la volontà supera i limiti dell'intelletto e costituisce lo spazio della libertà, con l'intelligenza artificiale, invece, molti processi cognitivi possono essere descritti come calcoli o simulazioni. Questo non implica ovviamente che la volontà umana scompaia, ma suggerisce che il libero arbitrio, così come concepito da Cartesio, deve essere ripensato in un contesto in cui le decisioni sono co-determinate da sistemi intelligenti e dati algoritmici.

La rottura si estende anche alla soggettività, Cartesio infatti distingue nettamente il soggetto pensante dal mondo esterno, oggi, l'interazione costante con le macchine intelligenti crea una soggettività distribuita in cui l'azione umana è mediata e influenzata da sistemi tecnologici complessi³¹. La responsabilità, quindi, non è più solamente individuale ma deve essere considerata in termini di ecosistema sociotecnico, in cui la tecnologia partecipa alla decisione pur senza essere agente morale autonoma.

L'uomo si trova così a confrontarsi con sistemi che ampliano le sue capacità, ma che al contempo limitano la sua autonomia nel senso tradizionale del termine.

2.4.3 Implicazioni per il libero arbitrio

L'analisi della continuità e della rottura tra il pensiero cartesiano e l'era dell'intelligenza artificiale evidenzia che la nozione di libero arbitrio va oggi interpretata in maniera più complessa. Se da un lato permane la rilevanza della

³⁰ Luciano Floridi, *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford University Press, 2014, cap. 6, pp.125-130.

³¹ Ivi, pp. 130-133.

distinzione tra intelletto e volontà, dall'altro il contesto tecnologico introduce vincoli e interazioni che rimodellano la libertà³².

Il soggetto, dunque, non è più isolato: le decisioni vengono prese in un ambiente in cui algoritmi e sistemi intelligenti possono anticipare, suggerire e condizionare le scelte umane. Ciò comporta sicuramente una redistribuzione della responsabilità. Nel modello cartesiano, la responsabilità era interamente attribuibile al soggetto agente, oggi essa va intesa come co-costruita all'interno di reti sociotecniche, dove l'uomo mantiene il ruolo di decisore finale ma in relazione costante con sistemi intelligenti che ne influenzano l'operato.

Un'ulteriore implicazione riguarda la definizione stessa di libertà: non più come facoltà assoluta e autonoma, ma come capacità di agire consapevolmente in un contesto complesso. La libertà diventa, in altre parole, una forma di libero e ritiro situato, in cui l'uomo decide tenendo conto delle condizioni imposte dalla tecnologia, dai dati e dalle reti di interazione³³.

Infine, l'analisi suggerisce che la riflessione sul libero arbitrio deve includere la considerazione della cooperazione tra intelligenza naturale e artificiale. Non si tratta solo di proteggere o riaffermare la volontà umana, ma di sviluppare strumenti concettuali e pratici per garantire che le decisioni umane restino informate, critiche e responsabili, anche quando mediate da sistemi intelligenti.

In sintesi, il libero arbitrio nell'era digitale non scompare, ma si trasforma: è necessario concepirlo come un esercizio critico e consapevole della volontà all'interno di un ecosistema complesso, in cui la responsabilità e la libertà si intrecciano con le capacità e i limiti della tecnologia.

³² Ivi, pp. 130-137.

³³ Ivi, pp. 135-138.

Capitolo 3- Il libero arbitrio nell'era dell'intelligenza artificiale

3.1 Cartesio e l'intelligenza artificiale

La diffusione delle tecnologie e dell'intelligenza artificiale costringe a ripensare i criteri che, nella prima modernità, hanno delimitato le possibilità dell'artificiale. In questo quadro, il contributo di Cartesio è significativo non perché anticipi l'IA contemporanea, ma perché elabora strumenti concettuali per distinguere l'imitazione meccanica dall'attività propriamente umana.

Un primo elemento significativo è la concezione della macchina artificiale come meccanismo capace di riprodurre l'esteriorità dei comportamenti senza accedere al loro significato. Nel *Traité de l'homme* Cartesio descrive il corpo umano come una struttura complessa, ma integralmente spiegabile attraverso rapporti meccanici e dinamici; spontaneità e vitalità non implicano alcun mistero naturale, bensì l'articolazione ordinata di movimenti e reazioni³⁴. Questa impostazione ha inaugurato un modo di pensare la corporeità come processo, aprendo la strada alla possibilità che un artefatto, se sufficientemente sofisticato, possa simulare funzioni considerate caratteristiche della vita.

Su questa base Cartesio formula due criteri destinati a distinguere l'uomo da qualunque costruzione artificiale: il primo è il criterio linguistico: nessuna macchina, per quanto dotata di organi perfetti, sarà in grado di «Disporre le parole in diversi modi per rispondere al senso di tutto ciò che si dice in sua presenza»³⁵.

Il secondo è il criterio dell'adattamento inventivo: un automa può eccellere in un compito specifico, ma fallirà quando si troverà di fronte a situazioni nuove, che richiedono iniziativa e flessibilità non programmabili³⁶. A differenza degli argomenti metafisici, questi due criteri hanno natura empirica: definiscono soglie comportamentali che, secondo Cartesio, nessuna macchina avrebbe potuto oltrepassare.

L'intelligenza artificiale contemporanea mette in questione entrambi i criteri senza necessariamente confutarli. Da un lato, i modelli linguistici di ultima generazione producono frasi variabili, coerenti e adattate al contesto, mostrando che l'uso flessibile del linguaggio non è una prerogativa esclusiva dell'umano. Dall'altro, il comportamento degli algoritmi di apprendimento automatico suggerisce una capacità di reagire a situazioni non previste dal progettista che, almeno in apparenza, sfida il secondo criterio cartesiano. Tuttavia, questa sfida riguarda il livello fenomenico delle

³⁴ René Descartes, *Traité de l'homme*, in *Oeuvres de Descartes*, a cura di Charles Adam e Paul Tannery, Paris, Vrin, 1996, pp. 99-103.

³⁵ René Descartes, *Lettere a Mersenne*, in *Correspondance*, in *Oeuvres de Descartes*, cit., vol. III, pp. 204-205.

³⁶ René Descartes, *Lettera al marchese di Newcastle*, in *Correspondance*, in *Oeuvres de Descartes*, cit., vol. III, pp. 303-304.

prestazioni, non quello della comprensione: che una macchina generi una risposta non implica che ne colga il senso, ed è proprio qui che Cartesio risulta attuale: il suo pensiero offre lo schema critico per distinguere tra efficacia operativa e attribuzione di significato, tra output funzionali e processi cognitivi.

Il valore attuale dell'approccio cartesiano sta dunque nell'obbligare a formulare con maggiore precisione che cosa intendiamo per pensiero e per linguaggio. L'IA non invalida Cartesio; piuttosto, impone di radicalizzarne la domanda: quali condizioni rendono un comportamento non solo efficace, ma comprensivo? Quali criteri consentono di distinguere un processo che appare intelligente da uno che lo è effettivamente? In un ecosistema tecnico che moltiplica le simulazioni di intelligenza, la soglia cartesiana tra imitazione e comprensione, lungi dall'essere un residuo storico, diventa un dispositivo critico indispensabile per interrogare natura, limiti e pretese dell'intelligenza artificiale.

3.2 La questione della libertà nell'età digitale

In questo paragrafo, l'obiettivo principale non è rispondere alla domanda se l'uomo sia ancora libero, ma piuttosto delineare il quadro in cui la libertà deve essere ripensata. La riflessione si concentra sull'interazione tra volontà umana e strumenti digitali, ponendo le basi per una comprensione più articolata della libertà nell'era digitale, senza cadere in semplici determinismi tecnologici o in astratte concezioni di autonomia pura.

L'ingresso delle tecnologie digitali nella quotidianità ha reso la questione della libertà personale un tema centrale del dibattito contemporaneo, non si tratta più soltanto di discutere la capacità dell'individuo di scegliere tra alternative morali o pratiche, come nella tradizione filosofica classica, ma anche di comprendere come le decisioni umane siano influenzate o mediate da sistemi algoritmici e intelligenze artificiali. La libertà, in questo contesto, non può più essere considerata esclusivamente come autonomia della volontà, ma deve essere analizzata alla luce di strutture di calcolo e modelli che operano in parallelo con l'agire umano.

Gli strumenti digitali intervengono nei processi decisionali attraverso la raccolta e l'analisi di dati personali, anticipando bisogni e suggerendo azioni, ciò introduce una tensione tra la percezione soggettiva della libertà e le condizioni oggettive imposte dalle tecnologie, l'individuo, infatti, può sentirsi libero, ma la sua scelta può essere già orientata da algoritmi che modellano comportamenti e preferenze. Questo fenomeno solleva molte questioni etiche e filosofiche, dal momento che il concetto di libertà non è più separabile dal contesto tecnologico in cui l'azione si svolge.

3.3 La decisione tra volontà e calcolo

Nell'era digitale, la decisione non è più soltanto un processo interiore dell'individuo, ma cresce dentro un contesto in cui le macchine partecipano sempre più spesso alla selezione delle alternative possibili. Gli algoritmi che regolano piattaforme e motori di ricerca, propongono scelte già pronte, orientando ciò che l'utente vede e di conseguenza anche ciò che considera rilevante³⁷. Questo significa che la volontà non sparisce, ma si trova immersa in un contesto che ne condiziona la scelta finale. Le intelligenze artificiali, infatti, non decidono come un soggetto umano: seguono regole, e modelli matematici, il loro obiettivo è individuare 'opzione più efficiente secondo dei criteri prestabiliti. In loro non c'è l'intenzionalità o la consapevolezza, ma solo calcolo³⁸. Tuttavia, proprio perché il calcolo risulta più veloce e semplice, le persone tendono ad affidargli una parte crescente delle proprie scelte quotidiane, dalle notizie che leggono ai percorsi che seguono, fino addirittura ai prodotti da acquistare. Questa delega parziale crea una situazione totalmente nuova: la decisione, infatti, diventa un processo condiviso e quasi frammentato tra volontà e calcolo. L'individuo rimane libero, ma molte delle sue scelte nascono in risposta a suggerimenti automatici che limitano o guidano le opzioni possibili, non si tratta quindi di una perdita totale della libertà, bensì di una trasformazione di essa: l'uomo decide, ma dentro una cornice definita dagli algoritmi³⁹. In questo scenario, la tensione centrale non è semplicemente tra "volontà umana" e "macchina", ma tra la libertà di scegliere e la forza delle previsioni che gli algoritmi producono su ciò che probabilmente sceglieremo, la sfida è quindi capire come mantenere uno spazio autentico di decisione personale in un mondo in cui molte scelte vengono anticipate prima ancora che emergano come possibilità nella mente del soggetto⁴⁰.

3.3.1 Intelligenza artificiale e notizie false

Come visto nel paragrafo precedente, l'IA è sempre più partecipe nelle nostre decisioni quotidiane, suggerendoci scelte e orientando i nostri comportamenti. Allo stesso modo, essa sta anche modificando il modo in cui percepiamo e riceviamo le informazioni, aumentando spesso il rischio di essere esposti a notizie false o fuorvianti. L'IA non si limita a facilitare l'accesso ai contenuti, ma può produrli autonomamente: testi, articoli, immagini e video vengono generati automaticamente sulla base di modelli predefiniti,

³⁷ Zuboff, S., *The Age of Surveillance Capitalism*, PublicAffairs, 2019, pp.274-282.

³⁸ Russel, S., Norving, P., *Artificial Intelligence: A Modern Approach*, 4th ed., Pearson, 2020, pp.34-36.

³⁹ Floridi, L., *The Logic of Information*, Oxford University Press, 2019, pp.118-122

⁴⁰ Agamben, G., *Il possibile e la necessità*, Einaudi, 2020, pp.45-49.

senza alcuna verifica della veridicità⁴¹. Questa caratteristica rende il controllo e la supervisione umana sempre più difficile, tradizionalmente l'uomo poteva valutare e filtrare le informazioni prima di diffonderle, con l'IA, invece, molte notizie vengono create o selezionate senza l'intervento umano diretto. Gli algoritmi privilegiano ciò che è sensazionalistico o coinvolgente, indipendentemente dalla loro accuratezza, perché il loro obiettivo è massimizzare le interazioni e la visibilità⁴².

In questo senso, la decisione di cosa credere o condividere non è più completamente nelle mani del soggetto: l'algoritmo condiziona la percezione delle alternative possibili, il problema però non riguarda solo la possibilità che vengano diffuse informazioni false, ma anche l'impatto che queste hanno sulla capacità dell'individuo di esercitare un giudizio libero. Se l'utente riceve costantemente contenuti manipolati o fuorvianti, la sua volontà si trova guidata da stimoli che egli stesso percepisce come neutri o oggettivi. Infine, l'opacità degli algoritmi ⁴³, complica ulteriormente la questione della responsabilità, non è sempre chiaro chi debba rispondere della diffusione di informazioni false: il programmatore, la piattaforma, o semplicemente l'utente che le condivide. Questa situazione evidenzia quanto sia urgente riflettere su come mantenere uno spazio di decisione autentico in un mondo dove la linea tra volontà e calcolo diventa sempre più sfumata⁴⁴.

3.4 Autonomia e determinismo nelle macchine intelligenti

Nel passaggio dal concetto umano di volontà deliberante alle decisioni prodotte dalle macchine intelligenti, emerge una tensione decisiva: ciò che chiamiamo "autonomia" dell'IA è compatibile con il carattere profondamente deterministico dei sistemi computazionali? Questa domanda diventa centrale in un contesto in cui, come visto nel paragrafo precedente, i sistemi artificiali non si limitano ad elaborare istruzioni fisse, ma trasformano la propria configurazione interna attraverso processi di apprendimento⁴⁵.

Se nei sistemi algoritmici tradizionali il determinismo è evidente, ogni output deriva da regole esplicite, il quadro cambia radicalmente con i sistemi di deep learning e machine learning, qui infatti la macchina sviluppa strutture interne che non sono

⁴¹ Floridi, L., Chiariatti, M., "GPT and the Limits of Artificial Intelligence", in *Minds and Machines*, vol. 30, 2020, pp. 99-104.

⁴² Vosoughi, S., Roy, D., Aral, S., "The Spread of True and False News Online", in *Science*, vol. 359, 2018, pp. 1146-1151.

⁴³ La difficoltà di comprendere perché un certo contenuto viene mostrato o generato.

⁴⁴ O'Neil, C., *Weapons of Math Destruction*, Crown, 2016, pp.57-61.

⁴⁵ J. Pearl, *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge 2009, pp. 45-52.

previste riga per riga dal programmatore. Nei sistemi di riconoscimento facciale, ad esempio, milioni di parametri vengono modificati attraverso l'esposizione ai dati, producendo risultati che non erano direttamente contenuti nelle regole iniziali e che difficilmente possono essere ricondotti a una singola istruzione⁴⁶. Questa complessità non segnala una libertà simile a quella umana, ma un'espansione delle possibilità operative entro un quadro matematico rigidamente definito.

Un caso particolarmente emblematico è quello degli algoritmi di raccomandazione utilizzati dalle grandi piattaforme digitali, come TikTok o Instagram, in cui il sistema non decide liberamente i contenuti da proporre ma li seleziona attraverso le statistiche di preferenza e di comportamento di milioni di utenti. La scelta è sempre vincolata da ciò che i dati mostrano, ma l'esito finale può assumere la forma di un comportamento imprevisto, quasi personalizzato che genera l'impressione di un'autonomia intenzionale⁴⁷.

È in questo spazio ibrido che studiosi come Luciano Floridi collocano la nozione di *agente artificiale*: non una macchina realmente libera, ma un'entità capace di auto-organizzazione funzionale entro precisi limiti computazionali. Questa autonomia funzionale rimane vincolata da due livelli di determinismo: il primo, inscritto nei modelli matematici e nelle architetture di apprendimento e il secondo, dipendente dalla qualità e dalla struttura dei dati utilizzati per addestrare il sistema⁴⁸.

Comprendere questa dinamica permette di definire con maggiore precisione che cosa significhi "autonomia" nelle macchine intelligenti: non una libertà in senso forte, ma la capacità di produrre azioni non direttamente determinate da un comando esterno, pur rimanendo iscritte in una cornice rigorosamente deterministica.

3.5 Libertà come emergenza o simulazione?

La riflessione sulla natura della libertà nell'era dell'intelligenza artificiale trova il punto cruciale nella domanda se ciò che appare come autonomia sia un fenomeno genuinamente emergente o soltanto una sofisticata simulazione del comportamento umano⁴⁹. Tale interrogativo nasce come naturale prosecuzione delle analisi precedenti: se, come discusso nel paragrafo precedente, i processi decisionali dell'IA oscillano tra calcolo deterministico e forme di adattamento che ne aumentano l'imprevedibilità, allora diventa necessario comprendere se questa "imprevedibilità" contenga davvero

⁴⁶ I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT press, Cambridge 2016, pp.85-90.

⁴⁷ C. O'Neil, *Weapons of Math Destruction*, Crown Books, New York 2016, pp. 121-142.

⁴⁸ L. Floridi, *The Ethics of Information*, Oxford University Press, Oxford 2013, pp. 83-91.

⁴⁹ J. Searle, *Mind, Language and Society*, New York, Basic Books, 1998, p. 112.

qualcosa di analogo alla libertà o se sia soltanto un'imitazione formalmente convincente⁵⁰.

L'ipotesi dell'emergenza sostiene che: raggiunto un certo livello di complessità, i sistemi artificiali producono proprietà non riducibili alla semplice somma dei loro componenti, analogamente a ciò che avviene nei sistemi biologici. Nel campo dell'IA, questa concezione è spesso richiamata a partire dagli sviluppi delle reti neurali profonde, le quali mostrano comportamenti non esplicitamente programmati, come la capacità di generare inferenze creative o di apprendere strategie inedite nei giochi complessi⁵¹. Per esempio, nel caso di *AlphaGo*, la mossa 37 contro Lee Sedol⁵² fu interpretata da molti come un atto quasi intuitivo, una scelta che sfuggiva allo schema tradizionale del puro calcolo enumerativo. Tale episodio ha alimentato il dibattito sull'eventualità che qualcosa di simile a una "spontaneità operativa" possa emergere anche in macchine prive di interiorità psicologica.

La posizione della simulazione insiste invece sul carattere puramente algoritmico di tali fenomeni. Secondo questa prospettiva, l'apparente libertà delle macchine non sarebbe che il risultato di modelli statistici estremamente complessi in grado di imitare le scelte umane senza alcuna forma di autodeterminazione autentica⁵³. L'IA in questa lettura, non sceglie ma calcola e non vuole ma ottimizza. È un apparato che restituisce uscite coerenti con ciò che ha appreso, ma non possiede intenzioni o capacità di attribuire significato alle proprie azioni. Ciò lo si può evincere nei casi in cui vengono generate risposte ambigue o logicamente contraddittorie, ciò dimostra che la macchina non discerne tra possibilità, ma semplicemente sceglie configurazioni probabilistiche sulla base di vari dati disponibili⁵⁴.

Il nodo teorico, dunque, riguarda il concetto stesso di libertà: se per essere liberi occorre essere consapevoli del proprio agire, allora l'emergenza computazionale non basterà mai a giustificare un'autonomia autentica della macchina. Se invece la libertà fosse definita come capacità di produrre comportamenti non totalmente determinati da una causa esterna e prevedibile, allora alcuni sistemi di IA contemporanei

⁵⁰ M. Boden, *Artificial Intelligence: A Very Short Introduction*, Oxford, Oxford University Press, 2018, p. 67.

⁵¹ Y. LeCun – Y. Bengio – G. Hinton, *Deep Learning*, in "Nature", vol. 521, 2015, pp. 436–444.

⁵² AlphaGo è un sistema di intelligenza artificiale sviluppato da DeepMind per giocare a Go, un gioco da tavolo strategico di origine cinese, grazie a reti neurali profonde e reinforcement learning ha raggiunto strategie non presenti nel gioco umano. In una partita contro il campione mondiale coreano Lee Sedol, la celebre "mossa 37" consistette in una scelta altamente improbabile secondo la teoria tradizionale del Go, interpretata come atto creativo poiché non imitava alcun modello umano noto e modificò in seguito la stessa teoria del gioco.

⁵³ N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford, Oxford University Press, 2014, pp. 82–83.

⁵⁴ G. Marcus, *Rebooting AI: Building Artificial Intelligence We Can Trust*, New York, Pantheon Books, 2019, p.59.

potrebbero sembrare in parte simili alla libertà umana. Resta comunque da sottolineare che anche i modelli più avanzati mancano di ciò che la filosofia ha storicamente considerato libertà, cioè un centro esperienziale, un io capace di riconoscersi come autore delle proprie azioni.

In questo senso, il dibattito tra emergenza e simulazione illumina le tensioni del nostro tempo: da un lato il fascino di un'intelligenza artificiale sempre più sofisticata, dall'altro la consapevolezza che la complessità tecnica non equivale automaticamente a un incremento dei diritti concettuali delle macchine.

La questione rimane quindi aperta, ma ciò che appare chiaro è che la somiglianza tra umano e artificiale non implica un'identità ontologica, e dunque non basta a garantire alla macchina un posto nel dominio della libertà così come viene comunemente concepita.

3.6 Conclusione: l'uomo di fronte alla macchina: chi è ancora libero?

Il percorso svolto in questo capitolo mostra che il confronto tra volontà umana e calcolo artificiale non produce una semplice equivalenza tra i due modelli decisionali, ma apre una frattura decisiva: la macchina calcola e l'uomo decide. Questa distinzione, benché resa instabile dallo sviluppo di sistemi capaci di apprendimento e autonomia operativa, rimane il nucleo della questione filosofica. Nessun algoritmo, per quanto sofisticato, può trasformare l'ottimizzazione in responsabilità o il successo operativo in libertà.

Nell'analizzare il rapporto tra determinazione algoritmica ed emergenza del comportamento delle macchine, abbiamo visto che l'imprevedibilità non coincide con la libertà: un sistema che sorprende non è un sistema che vuole⁵⁵. Questo punto è fondamentale in quanto svuota il campo da una delle illusioni più diffuse: credere che la somiglianza funzionale con l'umano sia già un'equivalenza ontologica. La libertà umana, al contrario, emerge nel momento in cui l'individuo riconosce i propri vincoli: biologici, psicologici e sociali, ma tuttavia può reinterpretarli e modificarli⁵⁶.

L'era dell'intelligenza artificiale non cancella questa possibilità, ma la rende più esigente, non perché la macchina sia più potente, ma perché diventa parte costitutiva dell'ambiente in cui la decisione avviene. La sfida, allora, diventa comprendere se l'uomo sia ancora capace di mantenere la distanza critica necessaria a distinguere ciò che vuole da ciò che gli viene suggerito.

La libertà qui, non è una proprietà interiore e nemmeno semplice atto volontaristico ma è un esercizio di vigilanza sul proprio modo di pensare, una forma di attenzione che

⁵⁵ M. Boden, *Artificial Intelligence: A Very Short Introduction*, Oxford University Press, Oxford 2018, pp.41-55.

⁵⁶ C. Taylor, *Sources of the Self*, Harvard University Press, Cambridge 1989, pp. 502-520.

nessun sistema può sostituire. Il punto non è competere con la macchina o con l'intelligenza artificiale, ma evitare di adottare come propri i loro criteri: efficienza, rapidità e ottimizzazione, lasciando che ridefiniscano silenziosamente ciò che consideriamo valido, desiderabile o vero⁵⁷.

In questo senso, l'uomo rimane libero solo se conserva la capacità di sottrarsi al flusso delle automazioni che lo circondano e sovrastano, questa sottrazione non deve essere vista come un atto eroico, ma come un atto cognitivo: comprendere come i processi decisionali vengano influenzati, quali alternative vengono escluse e quali possibilità vengono ampliate o ridotte a causa dell'interazione con la macchina.

Ciò che emerge, dunque, non è una risposta definitiva, ma una responsabilità crescente: essere liberi oggi significa imparare a distinguere l'assistenza dal condizionamento, il supporto della delega, l'autonomia dell'automatismo.

La libertà non scompare: cambia di forma e la sua nuova forma non è un dato, ma un compito.

⁵⁷ L. Floridi, *The Logic of Information*, Oxford University Press, Oxford 2019, pp. 201-225.

Capitolo 4- Etica, responsabilità e pensiero contemporaneo nell'era dell'intelligenza artificiale

4.1 Introduzione: la questione morale dopo l'automazione del pensiero

L'ingresso dell'intelligenza artificiale nella vita quotidiana e nella struttura decisionale delle società contemporanee ha prodotto uno scenario etico profondamente nuovo e complesso. Non si tratta più di affrontare dilemmi morali limitati all'ambito umano, ma di confrontarsi con entità capaci di elaborare informazioni, formulare soluzioni e talvolta persino anticipare comportamenti umani con una velocità e una precisione che l'uomo non può eguagliare⁵⁸. Questa trasformazione la necessità di ripensare la morale non solo come insieme di norme interiorizzate, ma come sistema capace di operare in un contesto in cui la responsabilità può essere distribuita tra soggetti biologici e sistemi artificiali.

L'automazione del pensiero, intesa come la possibilità che algoritmi complessi prendano decisioni su aspetti della vita economica, sociale e persino politica, solleva una domanda fondamentale: quale spazio resta alla valutazione morale umana quando il calcolo matematico diventa l'elemento predominante nella scelta?

La storia della filosofia morale mostra come i criteri di giusto e sbagliato siano stati tradizionalmente legati all'intenzionalità dell'agente. In un contesto in cui le decisioni possono essere prese da sistemi privi di coscienza, intenzioni o motivazioni emotive, l'intenzionalità, e quindi la valutazione morale convenzionale, diventa problematica⁵⁹. La questione non riguarda solo la responsabilità diretta delle macchine, ma anche il modo in cui esse influenzano la formazione della coscienza morale dell'uomo.

La facilità con cui un sistema può anticipare e indirizzare decisioni complesse rischia di instaurare una forma sottile di dipendenza, in cui la coscienza morale si appiattisce sulle soluzioni algoritmiche. Non si tratta di un giudizio tragico a priori, ma della constatazione che la centralità del calcolo, in assenza di un esercizio continuo della responsabilità etica, può ridurre lo spazio in cui l'uomo esercita realmente libertà e discernimento.

Un ulteriore aspetto riguarda la capacità dell'IA di operare su scala globale, le macchine intelligenti non sono vincolate dalle frontiere culturali o legali e possono applicare principi decisionali uniformi in contesti eterogenei. Questo porta a interrogativi di tipo cosmopolitico: è possibile definire una morale universale che si applichi in maniera coerente attraverso sistemi tecnologici che interagiscono con società diverse?

⁵⁸ Bostrom, N., *Superintelligence: Paths, Dangers, Strategies*, Oxford University Press, 2014, p.31

⁵⁹ Floridi, L., *The Ethics of Information*, Oxford University Press, 2013, p.47.

Oppure siamo destinati a confrontarci con una pluralità di etiche “programmate” che riflettono solo parzialmente la complessità dei valori umani?⁶⁰.

La questione morale, dunque, dopo l’automazione del pensiero, non riguarda soltanto ciò che le macchine possono fare, ma anche ciò che l’uomo diventa.

L’IA non sostituisce il pensiero umano in senso pieno, ma ne ridefinisce il contesto operativo, essa costringe a una revisione delle categorie etiche tradizionali, stimolando una riflessione su concetti quali responsabilità, intenzionalità, giustizia e libertà in un contesto in cui il confine tra agente e sistema diventa sempre più sfumato.

La questione morale nell’era dell’automazione del pensiero non è quindi un problema di nicchia, confinato al laboratorio di ricerca o alla programmazione tecnica, ma un nodo centrale per la sopravvivenza etica delle società contemporanee. Essa interroga il modo in cui definiamo il bene comune, come interpretiamo il ruolo della tecnologia nella vita quotidiana e fino a che punto possiamo affidarle compiti che fino a pochi decenni fa erano esclusivamente umani⁶¹.

4.2 Libertà e responsabilità: l’uomo davanti alle decisioni delle macchine

L’automazione cognitiva non ha eliminato la responsabilità umana: l’ha resa più ambigua, chi oggi delega scelte a sistemi intelligenti non si libera del peso decisionale, ma lo ricolloca in una zona grigia in cui la catena causale si allunga, si sporca e diventa più difficile da tracciare. Convincersi del contrario è solo un’illusione, una tecnologia opaca non è mai una giustificazione morale, ma solo una prova della nostra tendenza a evitare il conflitto con le conseguenze delle nostre stesse innovazioni⁶².

La difficoltà maggiore, nel rapporto tra libertà umana e decisioni delle macchine, riguarda il punto in cui si colloca la vera responsabilità, nella tradizione moderna, e specialmente in quella post-cartesiana, l’agire libero implicava un soggetto capace di assumere la causa della propria scelta come interiormente propria, distinta dall’ambiente e dagli strumenti di cui si serve.

L’intelligenza artificiale rompe linearità e proporzioni, una decisione è spesso il risultato di processi statistici, addestramento su dati non verificabili, ottimizzazioni che nessun operatore comprende fino in fondo. Ciò però non elimina la responsabilità

⁶⁰ Taddeo, M., Floridi, L., *The Ethics of Artificial Intelligence*, Springer, 2018, p.54.

⁶¹ Bynum, T., *Computer and Information Ethics*, Routledge, 2008, p.5.

⁶² P. Floridi, *The Ethics of Information*, Oxford University Press, Oxford 2013, p. 98.

del soggetto umano; semmai la rende più severa, perché l'opacità della macchina richiede un supplemento di vigilanza, non un alibi⁶³.

La responsabilità non si distribuisce automaticamente lungo la filiera tecnologica: va attribuita con rigore, distinguendo progettazione, implementazione, uso e supervisione. L'idea ingenua secondo cui "la colpa è dell'algoritmo" è una forma di deresponsabilizzazione che confonde determinazione tecnica e obbligo morale, un algoritmo infatti non sceglie ma ottimizza. Dunque, se una macchina produce un esito che danneggia un individuo, la domanda corretta non è "perché l'ha fatto?", ma "chi ha deciso di costruire il sistema in quel modo, o di utilizzarlo in quel contesto?".

Il fatto che una macchina sia più rapida, più precisa o più prevedibile di un umano non significa che debba sostituirlo nelle decisioni moralmente rilevanti, la tentazione tecnocratica è sempre la stessa: credere che l'efficienza equivalga alla giustizia, o che la coerenza algoritmica sia superiore alla fallibilità umana.

Le decisioni che incidono sulla vita delle persone, stabilire chi riceve un prestito, chi è considerato un rischio sociale, chi merita una cura prioritaria, chi ottiene un beneficio o un diritto, richiedono un livello di responsabilità che nessuna macchina può assumere, perché non può risponderne, se un sistema di scoring produce ingiustizia, non è "un problema tecnico", è una responsabilità morale che ricade su chi ha scelto di affidarsi a quel meccanismo senza controlli adeguati.

È necessario riconoscere che l'uso dell'intelligenza artificiale introduce un nuovo tipo di rischio morale: la delega strutturale.

L'uomo che delega troppo non è sollevato dalla responsabilità; è solo meno capace di riconoscere le conseguenze del proprio agire, la libertà non si esercita nel non decidere, ma nel vigilare sulle condizioni che rendono una decisione possibile. In questo senso l'automazione rende l'essere umano più vulnerabile: quando un sistema informatico decide al suo posto, l'uomo perde progressivamente il proprio "muscolo deliberativo", cioè la capacità di valutare, pesare, argomentare e assumere posizioni⁶⁴. Una libertà non esercitata non scompare all'istante, ma si atrofizza: il rischio è di ritrovarsi dipendenti da processi di cui non comprendiamo più la logica.

La responsabilità, allora, non può limitarsi al momento dell'errore: deve includere la progettazione istituzionale e culturale dell'uso dell'IA. Un sistema intelligente può essere un supporto, mai un sostituto, nelle scelte che comportano un giudizio di valore. Confondere i due livelli è il modo più rapido per perdere sia la responsabilità sia la

⁶³ L. Floridi, J. W. Sanders, "On the Morality of Artificial Agents", in *Minds and Machines*, vol. 14, 2004, pp. 349-379.

⁶⁴ N. Carr, *The Glass Cage: Automation and Us*, W. W. Norton, New York 2014, pp. 142-146.

libertà, lasciandoci con una macchina che decide e un uomo che firma, senza capire cosa ha avallato.

4.3 Prospettive etiche contemporanee

Le prospettive etiche contemporanee sull'intelligenza artificiale mostrano una tensione costante tra due esigenze fondamentali: da un lato comprendere l'IA come tecnologia progettata, controllabile e situata; dall'altro riconoscere che essa introduce forme di opacità, velocità decisionale e capacità operative che eccedono gli schemi tradizionali della responsabilità morale. Questa duplice tensione impone oggi una revisione profonda delle categorie classiche dell'etica normativa, mettendo in discussione la distinzione netta tra agente e strumento, tra autore dell'azione e mezzo tecnico impiegato⁶⁵.

Una prima linea interpretativa si concentra sulla necessità di responsabilizzare le strutture, più che i singoli individui. Studiosi come Luciano Floridi sostengono che nell'era digitale si debba adottare una "distributed morality", secondo cui la responsabilità è un fenomeno sistemico, condiviso tra progettisti, deployer, utenti e istituzioni ⁶⁶. Questo modello sposta l'attenzione dall'atto singolo al processo decisionale, suggerendo che l'etica dell'IA debba operare come una forma di "ingegneria del comportamento tecnologico", più che come giudizio retrospettivo di colpa.

Un secondo orientamento insiste invece sulla necessità di preservare la centralità del soggetto umano come criterio normativo ultimo, autori come Hubert Dreyfus e, più recentemente, Robert Sparrow hanno criticato le visioni che attribuiscono eccessiva autonomia morale ai sistemi artificiali, sostenendo che tali modelli rischiano di dissolvere il senso stesso della responsabilità, delegandola a entità non coscienti⁶⁷.

La questione non riguarda la sola capacità delle macchine di prendere decisioni, ma la possibilità di comprendere il significato morale delle proprie azioni, in assenza di intenzionalità, coscienza e comprensione fenomenologica, l'IA non può essere un "soggetto etico", ma solo un operatore funzionale i cui esiti devono essere sempre imputabili a esseri umani⁶⁸. Questa prospettiva mira a impedire che la crescente complessità algoritmica oscuri la necessità di un ancoraggio antropologico forte.

⁶⁵ Hans Jonas, *Il principio responsabilità*, Einaudi, Torino 1993, pp. 12-15.

⁶⁶ Luciano Floridi, *The Ethics of Information*, Oxford University Press, Oxford 2013, pp. 45-52.

⁶⁷ Hubert L. Dreyfus, *What Computers Still Can't Do*, MIT Press, Cambridge (MA) 1992, pp. 71-90.

⁶⁸ Robert Sparrow, "Can Machines be Morally Responsible?", in *Ethics and Information Technology*, vol. 7, n. 2, 2005, pp. 11-20.

Una terza posizione, più recente e maggiormente sensibile alle dinamiche reali dei sistemi intelligenti, propone invece un’“etica postumanistica della responsabilità condivisa”, pensatori come Rosi Braidotti e Donna Haraway invitano a superare la dicotomia rigida tra umano e non-umano, sostenendo che i sistemi di IA non vadano moralizzati, ma governati come attori ibridi con cui coabitiamo⁶⁹. L’obiettivo non è attribuire alle macchine una dignità morale, bensì riconoscere che esse sono già parte di un ecosistema di relazioni che produce effetti morali concreti: discriminazioni, esclusioni, redistribuzioni di potere, vulnerabilità nuove. Secondo questa prospettiva, l’etica dell’IA non riguarda ciò che “la macchina può fare”, ma ciò che noi diventiamo attraverso l’uso della macchina.

Accanto a queste correnti, si sviluppano approcci più operativi che tentano di integrare principi etici dentro i processi di sviluppo tecnologico, linee guida internazionali come quelle dell’High-Level Expert Group on AI della Commissione Europea suggeriscono un modello basato su quattro pilastri: trasparenza, equità, accountability e robustezza⁷⁰. Sebbene tali principi rappresentino un tentativo di formalizzare un’etica applicata, numerosi studiosi notano come essi rischiano di trasformarsi in un’etica “procedurale”, limitata a check-list che non affrontano le questioni strutturali di potere e disuguaglianza generate dalle nuove tecnologie. La difficoltà non è stabilire principi, ma tradurli in norme operative quando i sistemi intelligenti si comportano in modi non completamente prevedibili dai loro creatori.

Queste prospettive, diverse e talvolta incompatibili, convergono tuttavia su un punto decisivo: l’etica dell’intelligenza artificiale non è un’estensione dell’etica tradizionale, ma un ambito che costringe a ripensare le categorie stesse con cui abbiamo definito la responsabilità, l’agency e la libertà. Le macchine intelligenti non sono solo strumenti potenti, ma nuovi interlocutori del pensiero morale, che ci obbligano a riformulare la domanda: che cosa significa essere responsabili in un mondo in cui le nostre azioni si intrecciano con quelle di sistemi che apprendono, anticipano e agiscono in modi che non possiamo interamente prevedere?

4.3.1 Etica della progettazione delle macchine intelligenti: criteri, limiti e responsabilità degli sviluppatori

L’etica dell’intelligenza artificiale non può essere valutata solo a valle, osservando gli effetti delle decisioni che le macchine producono; va analizzata a monte, nel momento in cui i sistemi vengono progettati.

⁶⁹ Rosi Braidotti, *Il postumano*, DeriveApprodi, Roma 2014, pp. 89-102; Donna Haraway, *Staying with the Trouble*, Duke University Press, Durham 2016, pp. 23-25.

⁷⁰ European Commission, *High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI*, Brussels 2019, pp. 14-22.

Qui emerge il primo nodo critico: credere che l'IA sia un fenomeno “tecnico” e non una costruzione normativa è un'illusione comoda, che assolve i progettisti e deresponsabilizza le istituzioni.

Ogni algoritmo, infatti, non è solo una struttura matematica: è la materializzazione di valori umani incorporati nelle scelte di design, nei dataset selezionati, nei criteri di ottimizzazione e nelle assunzioni implicite che guidano lo sviluppo.

La retorica con cui spesso gli ingegneri si autoassolvono è “la macchina fa ciò per cui è ottimizzata”, un modo elegante per non guardare a chi ha deciso che cosa deve essere ottimizzato. Se un sistema massimizza l'engagement, questa non è una legge naturale, è una scelta economica, spesso opaca, che produce conseguenze morali precise

Se la progettazione diventa il luogo in cui si definiscono gli scopi della macchina, non si può più fingere che l'IA sia neutrale, è un dispositivo normativo che modella comportamenti, percezioni e anche la struttura delle opportunità dell'agente umano.

Da questo deriva la prima linea di responsabilità: gli sviluppatori non possono rifugiarsi nel mito della complessità tecnica essa non assolve, al massimo nasconde.

Per questo si è iniziato a parlare di etica incorporata (embedded ethics), non una vaga aspirazione, ma un insieme di procedure obbligatorie per rendere tracciabili, verificabili e discutibili le scelte di progettazione.

Una seconda questione riguarda la limitazione intenzionale delle capacità delle macchine, l'industria tecnologica opera spesso secondo un dogma implicito: ciò che può essere fatto tecnicamente, deve essere fatto, è un determinismo ingegneristico che confonde innovazione con progresso e che, se non viene contestato, trasforma l'etica in un freno a mano tirato⁷¹. In realtà, proprio perché le capacità tecniche sono cresciute in modo esplosivo, l'etica della progettazione deve includere la capacità di dire: questo non lo facciamo, non perché non possiamo, ma perché non dobbiamo.

Infine, c'è il nodo che molti evitano perché mina l'immagine eroica della Silicon Valley: la responsabilità non è solo degli ingegneri, ma delle istituzioni che finanziano, orientano e legittimano i modelli di IA. Per questo un'etica della progettazione efficace richiede non solo principi, ma vincoli giuridici, audit indipendenti, tracciamento delle decisioni di design, e soprattutto la possibilità democratica di contestare gli obiettivi delle tecnologie sviluppate.

4.3.2 Intelligenza artificiale e legislazione

L'avanzamento costante dell'IA nella nostra quotidianità, ha imposto una sfida non solo etica, ma anche giuridica, non basta più ragionare solo su ciò che è “giusto”

⁷¹ H. Jonas, Il principio responsabilità, Einaudi, Torino 1990, p. 34.

sviluppare o usare, ma serve un insieme di regole chiare, vincolanti e ripensate per una tecnologia che travalica i confini tradizionali delle responsabilità.

A livello europeo, il Regolamento AI Act (Regolamento UE 2024/1689) stabilisce un paradigma normativo basato sul rischio: non tutte le applicazioni dell'IA vengono trattate allo stesso modo, ma smistate secondo il potenziale di pericolo che possono generare. Ci sono sistemi considerati a basso rischio, altri a rischio limitato e quelli ad alto rischio, che richiedono obblighi più stringenti in termini di trasparenza, tracciabilità, documentazione, governance. Questo approccio bilancia protezione e innovazione, cercando di costruire un contesto in cui lo sviluppo dell'IA non sia anarchico, ma sotto controllo e con responsabilità ben definite per chi progetta e distribuisce questi sistemi.

In Italia, la recentissima legge 23 settembre 2025, n. 132 rappresenta un passo fondamentale nel recepire e integrare il modello europeo, ma con un tocco nazionale, viene infatti affermato un orientamento antropocentrico, cioè la centralità della persona umana e della supervisione umana nelle decisioni automatizzate.⁷²

Nella sua struttura, la legge prevede non solo principi generali di trasparenza, proporzionalità, non discriminazione, ma anche l'istituzione di una governance specifica: l'AgID (Agenzia per l'Italia Digitale) e l'ACN (Agenzia nazionale per la cybersecurity) sono chiamate a verificare il rispetto delle norme, mentre è prevista una strategia nazionale aggiornata ogni due anni.

Un aspetto particolarmente significativo della legge italiana riguarda la responsabilità: non è solo etica, ma penale, la legge introduce l'aggravante per reati commessi mediante sistemi di IA, e istituisce un nuovo reato specifico per la diffusione illecita di contenuti generati o alterati con l'IA (deepfake), con pene che possono arrivare fino a 5 anni di carcere.⁷³ Questo non è un dettaglio retorico, ma significa che lo Stato vuole imporre un vero vincolo giuridico sull'uso dell'IA, non semplici raccomandazioni.

Inoltre, per evitare che l'IA diventi un potente strumento di potere incontrollato, la legge stabilisce che nei settori sensibili come sanità, lavoro e giustizia, il ruolo umano deve restare centrale. In sanità e ricerca, ad esempio, l'IA può essere usata come supporto, ma non può prendere decisioni autonome che incidono su vite o scelte delicate. Questo vincolo di "supervisione umana effettiva" mette un freno all'idea che l'IA possa completamente sostituire l'uomo nei processi decisionali più critici.

⁷² The Law of September 23, 2025, no. 132 on Artificial Intelligence, LEXIA

⁷³ Italian Law on Artificial intelligence published in the Official Gazette, Andersen Italy

Sul piano dei diritti, la normativa italiana prevede anche la tutela dei dati personali in linea con il GDPR⁷⁴, richiedendo che le operazioni di trattamento legate all'IA siano trasparenti e che gli utenti siano informati in modo comprensibile sul funzionamento e sui rischi. Inoltre, per questioni di proprietà intellettuale, la legge stabilisce che le opere create con l'IA possono essere protette solo se riflettono un reale apporto creativo umano, non basta che siano generate automaticamente.

Tuttavia, nonostante i progressi, permangono delle criticità: la legge italiana delega molti aspetti al governo tramite decreti attuativi, e gran parte dell'efficacia reale dipenderà da come questi decreti saranno scritti e applicati, vi è anche il rischio che le imprese, soprattutto le piccole e medie imprese, trovino le nuove regole troppo onerose o complesse da implementare, specialmente in assenza di risorse adeguate e competenze legali o tecniche.

Infine, sul piano etico, la legislazione rappresenta un tassello necessario ma non sufficiente: una legge può imporre limiti, ma non garantisce che gli sviluppatori, le istituzioni e gli utilizzatori riflettano profondamente sul tipo di intelligenza artificiale che vogliamo costruire. Serve che il diritto si allinei con l'etica della progettazione, con audit indipendenti, processi di valutazione continua e partecipazione pubblica, senza questo, la regolamentazione rischia di restare una dichiarazione di buone intenzioni piuttosto che un freno reale a un uso irresponsabile dell'IA.

4.4 Etica e responsabilità in Cartesio

Il contributo di Cartesio al discorso etico non consiste in un sistema morale compiuto, ma in una struttura di responsabilità fondata sul rapporto tra intelletto e volontà.

Il punto essenziale è semplice e rigoroso, l'uomo è responsabile perché la volontà eccede ciò che l'intelletto comprende chiaramente, questa sproporzione non è un difetto, ma la condizione stessa della libertà, e proprio per questo impone un vincolo: assentire solo quando l'idea è chiara e distinta⁷⁵.

La conseguenza è una definizione di responsabilità radicalmente interna, l'errore non dipende dalle circostanze, ma dall'uso improprio della volontà, la sospensione del giudizio diventa così un dovere morale, non una debolezza.

In altre parole, l'etica cartesiana ruota attorno a un principio: la libertà non consiste nel decidere sempre, ma nel saper non decidere quando manca chiarezza.

⁷⁴ Acronimo di: General Data Protection Regulation, il Regolamento generale sulla protezione dei dati dell'Unione Europea, entrato in vigore il 25 maggio 2018. Il suo scopo è rafforzare la protezione dei dati personali e restituire agli individui il pieno controllo sui propri dati.

⁷⁵ R. Descartes, *Meditationes de prima philosophia*, in *Oeuvres*, Gallimard, Paris 1996, Meditazione IV, pp. 114–116.

Questa impostazione permette di distinguere tra libertà come potenza e libertà come rettitudine, per Cartesio, infatti, l'atto libero è tanto più perfetto quanto più aderisce alla verità: una volontà che si muove senza sufficiente evidenza non è più libera, ma arbitraria, è un criterio che rimette al centro la qualità del giudizio, e non la quantità delle opzioni.

Il valore attuale di questa prospettiva non sta nel comparare l'uomo alla macchina, ma nel sottolineare un punto che la discussione sull'IA spesso ignora, a responsabilità non riguarda gli effetti esterni delle azioni, ma la disciplina del giudizio che precede ogni decisione.

Cartesio ricorda che la libertà è inseparabile da un lavoro cognitivo, da una vigilanza sulle condizioni in cui si sceglie, e che nessuna tecnologia può sostituire questo lavoro senza ridurre la struttura stessa dell'agire umano. Per Cartesio, dunque, la vera etica non è un insieme di regole, ma l'esercizio rigoroso della libertà nel suo rapporto con la verità: un criterio che rimane esigente e attuale.

Capitolo 5 - Conclusioni: la libertà come responsabilità nell'era dell'intelligenza artificiale

La considerazione sul libero arbitrio pervade la storia della filosofia come un quesito che non si esaurisce, poiché concerne la stessa essenza dell'azione umana. Il confronto con il pensiero cartesiano e con i progressi attuali dell'intelligenza artificiale non vuole creare una netta separazione tra passato e presente, ma sottolineare come la questione della libertà continui a essere centrale anche in un contesto profondamente cambiato. In questa prospettiva, la tesi non appare come un traguardo finale, bensì come un tentativo di chiarire quali circostanze rendano ancora possibile, oggi, un autentico esercizio della libertà. Il contributo di Cartesio offre una struttura concettuale che resta rilevante poiché riconosce la complessità dell'azione umana, la differenza tra intelletto e volontà, e la loro diversa estensione; non cerca di limitare la libertà, ma di darle un fondamento: la volontà è vasta, forse illimitata; l'intelletto è finito. Pertanto, l'essere umano deve esercitare la libertà non come semplice forza, ma come abilità di valutare correttamente ciò che apprende, di ritardare l'approvazione quando manca di chiarezza e di guidare le proprie scelte in base alla verità riconosciuta. Questa configurazione mostra una dimensione morale specifica: la libertà si identifica con un impiego appropriato della volontà, e la responsabilità si manifesta come condizione intrinseca a tale impiego.

Quando la tecnologia contemporanea introduce sistemi capaci di apprendere, valutare, prevedere e decidere, non si tratta solo di verificare se le macchine siano "libere" o meno, ma di capire come l'esistenza di questi sistemi cambi il modo in cui l'essere umano esercita la propria libertà.

L'intelligenza artificiale non agisce con intenti, non mira a obiettivi e non pone questioni morali; opera tramite correlazioni e ottimizzazioni statistiche, producendo decisioni che possono influenzare concretamente la vita delle persone e delle comunità, la sua "autonomia" è pratica, non etica.

Per questo motivo, la questione essenziale non concerne ciò che l'IA "è in grado" di fare, ma ciò che l'individuo sceglie di fare, fidandosi o opponendosi ai suoi processi decisionali.

La libertà nell'epoca dell'intelligenza artificiale prende così una nuova dimensione, dove la facoltà di giudizio diviene fondamentale.

L'individuo non è più solo chi seleziona tra varie opzioni, ma chi deve riconoscere quali scelte siano veramente sue e quali siano invece influenzate, orientate o addirittura suggerite dai sistemi intelligenti; in questo scenario, la responsabilità non si presenta come un elemento supplementare, ma come la condizione fondamentale che consente alla libertà di preservare la propria autenticità: essere liberi implica prendersi il peso

della decisione, anche quando strumenti esterni cercano di agevolare, semplificare o guidare tale processo.

La dimensione etica della libertà, che Cartesio associa al giusto equilibrio tra volontà e intelletto, si espande in un contesto in cui partecipano attori non umani; la tecnologia non annulla il ruolo dell'agente umano, ma lo trasforma.

Essa presenta nuove opportunità di intervento, nuovi strumenti conoscitivi e nuove abilità operative; contemporaneamente introduce rischi di delega non critica e margini d'errore differenti. In tal senso, la libertà necessita di una sorveglianza più accurata: non è sufficiente decidere, è necessario afferrare le circostanze in cui avviene la decisione. L'atto libero non è un gesto involontario, bensì un procedimento che richiede consapevolezza, riflessione e attenzione alle ripercussioni.

Un altro aspetto significativo concerne il significato di "autonomia" nell'era digitale; se nelle Meditazioni l'autonomia è la facoltà umana di controllare le proprie inclinazioni tramite il giudizio, oggi essa comporta anche la facoltà di gestire gli strumenti che mediano il nostro legame con il mondo. La libertà non si limita all'opposizione tra naturale e artificiale, ma coincide con la capacità dell'individuo di mantenere la propria responsabilità in un contesto tecnologico complesso; l'agente libero è quindi chi integra la tecnologia senza perdere la propria autoconoscenza.

L'intera traiettoria di questa tesi porta a una conclusione morale chiara: la libertà nell'epoca dell'intelligenza artificiale non deve essere vista come un semplice dato di fatto, ma come un impegno che implica discernimento, consapevolezza e responsabilità, un impegno che non mira a controllare completamente la tecnologia, ma a capire come questa influisca sulle condizioni del nostro operare.

Il pensiero cartesiano, sebbene appartenga a un diverso contesto storico, fornisce una cornice concettuale che continua a supportare l'interpretazione di questa sfida: la libertà non deriva dal potere di scegliere, ma dalla qualità del giudizio che guida la scelta. È in questo rapporto fra intenzione e consapevolezza, fra scelta e valutazione, che risiede la dignità dell'azione umana; in definitiva, la libertà attuale non è in contrasto con la tecnologia, ma si esprime nel modo in cui l'uomo interagisce con essa. L'intelligenza artificiale non elimina la libertà, ma la sfida; non elimina la responsabilità, ma la rende più rigorosa. L'essere umano resta libero non evitando l'IA, ma continuando a riflettere sulle proprie intenzioni, le proprie decisioni e le proprie responsabilità.

In conclusione, quindi, ha un valore totalmente etico: essere liberi implica adottare un approccio attento e riflessivo, capace di incorporare la tecnologia senza perdere di vista l'importanza dell'agire morale. La libertà si presenta come un impegno che si rinnova quotidianamente, come un esercizio di valutazione e responsabilità che non può essere

trasferito, e come la dimensione più autentica dell'essere umano nell'era delle macchine intelligenti.

Bibliografia

Agamben, G., *Il possibile e la necessità*, Einaudi, 2020.

Boden M., *Artificial Intelligence: A Very Short Introduction*, Oxford, Oxford University Press, 2018.

Boole, G., *An Investigation of the Laws of Thought*, Dover, 1958.

Bostrom N., *Superintelligence: Paths, Dangers, Strategies*, Oxford, Oxford University Press, 2014.

Braidotti R., *Il postumano*, DeriveApprodi, Roma 2014.

Bynum, T., *Computer and Informaton Ethics*, Routledge, 2008.

Carr, N., *The Glass Cage: Automation and Us*, W. W. Norton, New York 2014.

Descartes, R., *Discourse de la Mèthode* (1637), Mondadori, Milano 1993.

Descartes, R., *Epistolae ad Elisabet*, 6 ottobre 1645, IV, 266.

Descartes R., *Lettera al marchese di Newcastle*, in *Correspondance*, in *Oeuvres de Descartes*, cit., vol. III.

Descartes R., *Lettere a Mersenne*, in *Correspondance*, in *Oeuvres de Descartes*, cit., vol. III.

Descartes R., *Meditationes de prima philosophia*, in *Oeuvres*, Gallimard, Paris 1996,
Descartes R., *Meditations on First Philosophy*, in J. Cottingham, Cambridge University Press, 1986, *Meditazione II*.

Descartes R., *Meditazioni metafisiche*, trad. Di E. Grain, Laterza, 1997.

Descartes, R., *Prefazione ai principi di filosofia*.

Descartes, R., *Principia Philosophiae*, vol. VIII.

Descartes R., *Trattato dell'uomo*, 1622.

Dreyfus Hubert L., *What Computers Still Can't Do*, MIT Press, Cambridge (MA) 1992.

European Commission, *High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI*, Brussels 2019.

Floridi L., *The Philosophy of Information*, Oxford University Press, 2011.

Floridi L., *The Ethics of Information*, Oxford University Press, Oxford 2013.

Floridi L., *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford University Press, 2014.

- Floridi, L., *The Logic of Information*, Oxford University Press, 2019.
- Floridi, L., Chiariatti, M., “GPT and thr Limits of Artificial Intelligence”, in *Minds and Machines*, vol. 30, 2020.
- Floridi L., J. W. Sanders, “On the Morality of Artificial Agents”, in *Minds and Machines*, vol. 14, 2004.
- D. Garber, *Descartes’ Metaphysical Physics*, Chicago, University of Chicago Press, 1992.
- Gaukroger S., *Descartes. An intellectual Biography*, Oxford, Clarendon Press, 1995.
- Goodfellow I., Y. Bengio, A Courville, *Deep Learning*, MIT press, Cambridge 2016.
- Haraway D., *Staying with the Trouble*, Duke University Press, Durham 2016.
- Hans Jonas, *Il principio responsabilità*, Einaudi, Torino 1993.
- LEXIA, The Law of September 23, 2025, no. 132 on Artificial Intelligence.
- LeCun, Y. – Y. Bengio – G. Hinton, *Deep Learning*, in “Nature”, vol. 521, 2015.
- Luger, G. F., *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, Pearson, 2020.
- Marcus G., *Rebooting AI: Building Artificial Intelligence We Can Trust*, New York, Pantheon Books, 2019.
- Mori G., *Cartesio*, Carocci editore, 2010.
- O’Neil C., *Weapons of Math Destruction*, Crown Books, New York 2016.
- Pearl J., *Casuality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge 2009.
- Poole, D., Mackworth, A., & Goebel, R. *Computational Intelligence: A Logical Approach*, Oxford University Press, 2017.
- Riskin J., *The Restless Clock: A History of the Centuries-Long Argument over What Makes Living Things Tick*, University of Chicago Press, 2016.
- Russel, S., & Norving, P. *Artificial Intelligence: A Modern Approach*, IV ed., Pearson, 2021.
- Searle J., *Mind, Language and Society*, New York, Basic Books, 1998.
- Sparrow R., “Can Machines be Morally Responsible?”, in *Ethics and Information Technology*, vol. 7, n. 2, 2005.
- Taddeo, M., Floridi, L., *The Ethics of Artificial Intelligence*, Springer, 2018.

Taylor C., *Sources of the Self*, Harvard University Press, Cambridge 1989.

Turing, A. M., “*On Computable Numbers, with an Application to the Entscheidungsproblem*”. Proceedings of the London Mathematical Society, Series 2, vol 42. 1936.

Vosoughi, S., Roy, D., Aral, S., “*The Spread of True and False News Online*”, in Science, vol. 359, 2018.

Wiener, N., *Cybernetics: Or Control and Communication in the Animal and the Machine*, MIT Press, 1948.

Zuboff, S., *The Age of Surveillance Capitalism*, PublicAffairs, 2019.

Dichiaro che, in alcune fasi di revisione linguistica e di organizzazione preliminare delle informazioni, ho utilizzato un sistema di intelligenza artificiale generativa (*ChatGPT, OpenAI*) come strumento di supporto. Tale utilizzo ha avuto esclusivamente funzione ausiliaria e non ha sostituito il lavoro di ricerca, analisi delle fonti e interpretazione critica da me svolto.