

Elvio Ancona\*

*Quale antropologia per l'IA? Prospettive di ricerca*

ABSTRACT

This essay intends to outline a philosophical-juridical research perspective on the anthropological foundation of the design of Artificial Intelligence (hereinafter AI). In particular, it aims to indicate which anthropological conception should guide the solution of the legal problems connected to the instrumental relationships that are established between AI and human subjects. Pursuing this objective, after having examined the main anthropological conceptions of our time, consequentialism, deontologism and natural law theory, it is argued that a personalist anthropology, supported by a philosophical-dialectical justification, is the one that allows us to look at the human being always also as an end and never simply as a means of intelligent systems. In the conclusions, with the aim of promoting the protection of human dignity, it is proposed to reactivate a philosophical anthropology, and more specifically to rediscover the importance of the philosophy of law, especially with reference to the theory of fundamental rights and the conditions of their protection.

KEYWORDS

Artificial Intelligence, Anthropology, Ethics, Personalism, Person

INDICE

1. Introduzione – 2. L'inversione della relazione mezzo-fine – 3. La necessità di un'indagine antropologica – 4. Quale antropologia? – 5. Un approccio (apparentemente) neutrale: il paradigma dei principi – 6. La prospettiva personalista – 7. La sfida riduzionista – 8. Una diversità irriducibile – 9. Quale personalismo? – 10. Conclusioni: il compito della filosofia.

1. Introduzione

Il presente saggio intende delineare una prospettiva di ricerca filosofico-giuridica sulla fondazione antropologica del *design* dell'Intelligenza Artificiale (di seguito IA). Spesso si parla infatti di etica dell'IA, oppure si considerano i problemi giuridici che l'impiego dell'IA pone quando i soggetti umani ne sono a diverso titolo coinvolti, dai programmatori a coloro che ne subiscono gli effetti o ne conseguono vantaggi. Capita raramente invece di vedere trattata e discussa la concezione antropologica fondamentale che di fatto sorregge tale impiego o quella di cui esso dovrebbe auspicabilmente tener conto. Questa questione per lo più non viene affrontata, in quanto

---

\* Professore associato presso l'Università degli Studi di Udine. E-mail: [elvio.ancona@uniud.it](mailto:elvio.ancona@uniud.it)

ritenuta filosoficamente insolubile e fomite di interminabili e inconcludenti dibattiti. Si preferisce occuparsi delle conseguenze, ritenendo sufficiente, quanto alle premesse, riferirsi ad un generico umanesimo, senza rendersi conto del fatto che esso risulterà inevitabilmente insufficiente quando si tratterà di confrontarsi con le visioni riduzioniste, transumaniste e postumaniste, dotate di un impianto ideologico dalla portata sistematica e onnicomprensiva.

Il presente saggio si propone di affrontare la questione antropologica esaminandone un aspetto specificamente giuridico. In particolare, esso si propone di indicare quale concezione antropologica dovrebbe orientare la soluzione delle problematiche giuridiche connesse alle relazioni strumentali che si instaurano tra l'IA e i soggetti umani.

## 2. L'inversione della relazione mezzo-fine

L'evoluzione dell'informatica, soprattutto quella dell'IA, pone il rapporto tra uomo e tecnica davanti ad una nuova sfida. Come tutti gli strumenti, anche il sistema informatico, infatti, replicando sul piano tecnologico la dialettica hegeliana del servo e del padrone<sup>1</sup>, può asservire colui che dovrebbe servire<sup>2</sup>, con tutte le conseguenze che questo comporta, per esempio in tema di violazione di diritti umani fondamentali. Ma in questo caso sembra che un simile esito nefasto possa dipendere, non da un impiego improprio del mezzo tecnologico da parte dell'uomo, ma dalla sua capacità di divenire fine a sé stesso. E con ciò non ci si riferisce appena al caso estremo della 'singolarità'<sup>3</sup>, quanto piuttosto alla logica intrinseca che lo innerva, la logica di un autosviluppo incessante e soprattutto in-definito. Tale logica, in sé tipica della tecnica moderna<sup>4</sup>, in quanto fatta propria dai sistemi informatici e dall'IA<sup>5</sup>, trova ormai riscontro nella capacità di quest'ultima di agire come soggetto decisore<sup>6</sup>, autonomo<sup>7</sup>, determinante se non altro obiettivi intermedi o

<sup>1</sup> Cfr. Hegel 1973 [1807]: 153 ss.

<sup>2</sup> Lo paventa, con riferimento al dispositivo robotico, Fabris 2018: 50.

<sup>3</sup> Si vedano in proposito, almeno, i fondamentali saggi di Vinge 1993 e Kurzweil 2005.

<sup>4</sup> Lo esprime efficacemente, citando Severino (2008: 37-66), Umberto Galimberti: "L'età della tecnica prende avvio quando l'uso della tecnica non ha più in vista alcuna finalità (neppure il profitto), ma solo il proprio potenziamento. E ciò accade quando appare chiaro che il raggiungimento di qualsiasi fine è subordinato alla disponibilità tecnica, per cui il potenziamento di questa disponibilità finisce con il rappresentare l'unico vero fine" (Galimberti 1999: 399).

<sup>5</sup> Lo evidenzia, questa volta con riferimento ad Heidegger (1988 [1984]), Adriano Fabris, notando come per il filosofo tedesco la cibernetica, la "nuova scienza" che farà da incubatrice dell'IA, costituisca "il momento più avanzato, l'esito più evidente di quel dominio della tecnica in cui sfocia l'intera metafisica occidentale" e in quanto tale finisca per ridurre l'attività umana a qualcosa di funzionale e controllabile dalla macchina: "L'uomo stesso diviene 'qualcosa di pianificato', cioè di controllabile e, posto che una tale riduzione non sia possibile, viene messo fra parentesi quale 'fattore di disturbo' nel calcolo cibernetico" (Fabris 1988: 10-11).

<sup>6</sup> Cfr. Simoncini 2019: 69: "Il fatto è che oggi la tecnologia (o per utilizzare la terminologia che propongo, il potere *cibernetico*) non è più soltanto un 'mezzo' per realizzare un corso di azioni deciso da un soggetto agente umano, ma, sempre più spesso, è *essa stessa a prendere decisioni rilevanti per la persona umana e la sua libertà*. Alla macchina non si chiede di realizzare ciò che un soggetto ha deciso, ma le si chiede di decidere, autonomamente".

<sup>7</sup> Per il significato che assume l'"autonomia" nel presente contesto si rinvia ai *consideranda* contenuti nella *Carta della robotica* approvata dal Parlamento Europeo nel febbraio 2017: "l'autonomia di un robot può essere definita come la capacità di prendere decisioni e metterle in atto nel mondo esterno, indipendentemente da un controllo o influenza esterna; [...] tale autonomia è di natura puramente tecnologica e il suo livello dipende dal grado di complessità con cui è stata progettata l'interazione di un robot con l'ambiente" (P8\_TA [2017] 0051). Sulle diverse concezioni e tipologie di "autonomia" delle macchine dotate di IA, cfr. Casadei e Pietropaoli 2021: 228-9. Sulla qualificazione dell'"autonomia" dell'IA come «autonomia funzionale», cfr. Quintarelli (cur.) 2020: 78. Per alcune riserve, infine, sull'applicabilità del termine alle macchine in generale, si vedano le osservazioni di Palazzani 2020: 78-79.

individuante finalità subordinate<sup>8</sup>, e quindi ultimamente in grado di strumentalizzare, almeno entro questi limiti, i soggetti di cui dovrebbe essere lo strumento<sup>9</sup>. Vale la pena di riportare al riguardo l'efficace sintesi di Rosaria Piroso:

“Le teorie critiche sulla tecnoregolazione prefigurano l'avvento di ‘master algorithms’ capaci di orientare i processi decisionali senza l'*input* o il controllo umano verso un'‘algorithmic governance’ o un'‘algocracy’ che, nella dimensione attuale, si manifesta nel fatto che gli algoritmi sono crescentemente utilizzati, a prescindere da una determinazione personale volontaria e più o meno consapevole, per persuadere, presentare parametri di valutazione, guidare e al fine di provocare, controllare, manipolare e coartare il comportamento umano. Emerge, dunque, una normatività diversa da quella giuridica in cui l'algoritmo si estrinseca non soltanto nella sua funzione ‘predittiva’, ma anche, come mostrano gli strumenti computazionali, nella sua portata prescrittiva e, di conseguenza, ‘veridittiva’. L'impiego di reazioni, interazioni e comportamenti umani nel funzionamento algoritmico, infatti, capovolge la relazione ‘utilizzatore-utilizzato’ rappresentando un tassello fondamentale nella complessa dinamica tecnoregolativa”<sup>10</sup>.

Le conseguenze in termini di violazione della dignità e di diritti umani fondamentali cominciano già a manifestarsi chiaramente<sup>11</sup>, aggravate dall'opacità dei processi decisori dell'IA<sup>12</sup>. Scrive in proposito Laura Palazzani: “L'aumento del potere tecnologico sull'umano apre nuove opportunità, ma anche sfide, rendendo l'uomo sempre più vulnerabile, esponendolo a nuove forme di strumentalizzazioni e discriminazioni”<sup>13</sup>. Si parla del resto di ‘nuovi diritti’ che garantiscano una collaborazione uomo-macchina, che non sia sostitutiva o asservente nei confronti dell'operare umano<sup>14</sup>.

Si manifesta dunque, soprattutto da un punto di vista giuridico, la necessità di un'indagine che consideri l'interazione tra sistemi artificiali ed esseri umani nella prospettiva della relazione mezzo-fine. Un'indagine siffatta, però, non può limitarsi solamente agli aspetti tecnologici, giuridici e nemmeno filosofici di una corretta impostazione di questa relazione e delle conseguenze tragiche della sua inversione. Deve essere innanzitutto un'indagine di orientamento antropologico.

### 3. La necessità di un'indagine antropologica

È chiaro, infatti, che il soggetto umano è interessato non solo in quanto programmatore o fruitore delle operazioni svolte dal sistema informatico, ma innanzitutto in quanto, come ogni artefatto, anche questo tipo di sistema, anche l'IA, reca in sé quella che abbiamo chiamato la “traccia indelebile dell'origine”<sup>15</sup>. Per quanto esso sia in grado di operare autonomamente, non può togliersi di dosso il ‘marchio di fabbrica’, la dipendenza dalla mente dell'artefice, colui che gli ha dato vita

<sup>8</sup> Il *modus operandi* dell'IA può essere ben illustrato da quello dei sistemi robotici, come i veicoli a guida autonoma e le armi autonome. Si veda al riguardo l'ampia e variegata casistica esaminata in Tamburrini 2020.

<sup>9</sup> Su questo rischio, cfr. Head 2014.

<sup>10</sup> Piroso 2021: 280-281. Per un approfondimento sul concetto di ‘algocracy’, si veda Danaher et al. 2017.

<sup>11</sup> Si pensi al caso dell'algoritmo dei riders (cfr. Tribunale di Bologna, n.r.g. 2949/2019) o degli insegnanti (cfr. Tar Lazio Sez. III bis n. 3769/2017; n. 9224-0230/ 2018) o all'ancor più celebre caso statunitense dell'algoritmo “Compass” (cfr. State Vs. Loomis, 881 N.W.2d 749 [Wis. 2016]).

<sup>12</sup> Ci si riferisce al fenomeno noto come ‘black box’, per il quale si rimanda, *in primis*, a Knight 2017.

<sup>13</sup> Palazzani 2020: 68.

<sup>14</sup> Cfr. *ivi*: 75 ss.

<sup>15</sup> Cfr. Ancona 2019: 83.

conferendogli il proprio patrimonio genetico, non può liberarsi dal vincolo rappresentato da tale soggettività<sup>16</sup>, non tanto come limite operativo, quanto in relazione alle finalità sue proprie. Questo, se non altro, è esemplificato dalla dipendenza, mediata o immediata, dell'IA dal costruttore umano; così come del resto avviene per qualsiasi altro prodotto tecnologico:

“Sarebbe un grave errore separare il lavoro produttivo dalle convinzioni morali, dai pregiudizi e dalle peculiarità che contraddistinguono la visione del mondo di chi lo svolge. Tra le due sfere – tra ciò che si pensa e ciò che si fa – esiste una relazione complessa e la commistione di caratteri tecnici e etico-sociali è una condizione ineliminabile della produzione di strumenti. Cosa ancora più vera nel caso dell'Intelligenza Artificiale”<sup>17</sup>.

Ma allora diventa indispensabile tornare a interrogarsi sull'essere umano, su ciò che lo caratterizza e lo determina ontologicamente, ma anche su come deve concepirsi per poter continuare a usare le macchine, questa specie di macchine soprattutto, senza divenirne schiavo. “Una cosa è certa: - scrive Remo Bodei - che più si impone l'automatismo, più siamo obbligati a non essere noi automatici e a sviluppare la coscienza accanto alla conoscenza”<sup>18</sup>.

#### 4. Quale antropologia?

Non possiamo a questo punto evitare un cenno alle concezioni antropologiche su cui si basano le principali teorie etiche contemporanee (utilitarismo, deontologismo e giusnaturalismo, per limitarsi alle più importanti)<sup>19</sup> e al differente impatto che possono avere sulla strutturazione dell'interazione tra soggetti artificiali e umani dal punto di vista della relazione mezzo-fine.

Limitiamoci a considerare le principali, a titolo esemplificativo. Tutte sono apparse gravate da importanti criticità e sono state ripensate in modo da poterle superare<sup>20</sup>.

L'*utilitarismo* va incontro ad un'obiezione ormai classica: la più grande felicità del maggior numero di persone può realizzarsi a scapito dell'infelicità di alcuni, dei loro diritti e istanze. A tale obiezione risponde quello sviluppo dell'utilitarismo meglio noto come consequenzialismo: esso, infatti, distingue scopi che sono beni in senso empirico (riguardanti il benessere fisico e materiale) e scopi che sono beni in senso propriamente razionale e morale. Il consequenzialismo tende così a coincidere con il teleologismo, che supera l'obiezione classica nella misura in cui riconosce fini che vanno perseguiti indipendentemente dai vantaggi personali o collettivi che si possono conseguire.

<sup>16</sup> Remo Bodei (cfr. Bodei 2019: 299-300, n. 6) ricorda in proposito i versi del coro delle macchine di Kipling: “Anche se il nostro fumo adesso nasconde il cielo ai vostri occhi / Presto svanirà e le stelle torneranno a brillare / Perché con tutta la nostra potenza e il peso e la grandezza che abbiamo / Noi non siamo altro che figlie del vostro sapere!” (traduzione dal poema *The Secret of the Machines*, raccolto in Eliot 1961: 293-294).

<sup>17</sup> Quintarelli (cur.) 2020: 95.

<sup>18</sup> Bodei 2019: 329.

<sup>19</sup> Per un quadro sintetico, cfr. Fabris 2018: 27-28. Secondo una formulazione più diffusa le tre teorie normative fondamentali sono il consequenzialismo, il deontologismo e l'etica delle virtù. Come presto vedremo, le due classificazioni non sono affatto incompatibili, potendosi ritenere consequenzialismo ed etica delle virtù sviluppi o varianti rispettivamente dell'utilitarismo e del giusnaturalismo. Per maggiori approfondimenti sulle concezioni di questo secondo gruppo, cfr. Baron, Pettit, and Slote, 1997. Per una presentazione più ampia e articolata delle molteplici teorie impegnate nel dibattito contemporaneo sull'etica normativa, si rinvia infine ai saggi raccolti in Singer (ed.) 2006: xi-xiii, 159 ss.

<sup>20</sup> Si vedano in proposito, soprattutto in riferimento a quanto segue nel testo, le considerazioni sviluppate in Da Re 2010: 83-106.

Il *deontologismo*, stabilendo il primato della legge e del dovere sul bene, va incontro alle obiezioni del formalismo e, in situazioni morali particolarmente complesse, del rigorismo. A tali obiezioni rispondono le teorie che hanno cercato di riproporre l'impianto deontologico sostituendo a leggi e doveri principi personalistici, come il rispetto della dignità dell'uomo e dei suoi diritti fondamentali.

Il *giusnaturalismo* è spesso visto come richiamantesi a un codice di leggi generali e immutabili, astratto dall'esperienza e pregiudicato da *bias* metafisici o religiosi. Questa è però solo una particolare specie di giusnaturalismo, di impronta essenzialmente normativista. Il giusnaturalismo può anche essere espresso come una teoria dei beni umani e del ragionamento pratico per conseguirli, ovvero come una teoria delle virtù e dell'uomo buono, o ancora come una concezione personalista, basata sul riconoscimento delle caratteristiche di razionalità e libertà dell'essere umano, e della sua capacità di perseguire autonomamente finalità proprie.

Quale tra queste concezioni potrebbe meglio supportare una prospettiva orientata al recupero della centralità del soggetto umano nell'impiego delle tecnologie informatiche e dell'IA? Di quale antropologia abbiamo bisogno? Lasciamo per il momento aperta la domanda, limitandoci a sottolineare la sua essenzialità, e la sua imprescindibilità.

## 5. Un approccio (apparentemente) neutrale: il paradigma dei principi

Proprio la riluttanza a impegnarsi con un preciso orientamento antropologico può aver indotto alcuni autori a propendere per un modello neutro, basato su alcuni principi etici fondamentali, per lo più utilizzati nelle numerose dichiarazioni per lo sviluppo di un'IA socialmente vantaggiosa. Emblematico è il caso di Luciano Floridi che nella sua recentissima *Etica dell'intelligenza artificiale*<sup>21</sup> illustra cinque principi comuni ai sei documenti di maggior profilo da poco elaborati dagli esperti del settore<sup>22</sup>: segnatamente, i principi di beneficenza, non maleficenza, autonomia, giustizia ed esplicabilità<sup>23</sup>.

Anche quest'approccio presenta tuttavia notevoli criticità. In primo luogo, come rileva lo stesso Floridi, questi principi sono ricavati da documenti elaborati "all'interno delle democrazie liberali occidentali", mentre vi sono regioni e culture che non vi sono "attualmente rappresentate o sono sottorappresentate"<sup>24</sup>. In secondo luogo, - è ancora Floridi che lo sottolinea - la loro universalità può essere messa in discussione anche da un altro punto di vista: "il significato di fondo di ciascuno dei principi è contestato, in quanto termini simili sono spesso usati per significare cose diverse"<sup>25</sup>. Infine, come spesso capita nel diritto costituzionale, si possono determinare situazioni in cui due o più principi si trovano a collidere tra loro, facendo insorgere, in relazione ai rispettivi assetti

<sup>21</sup> Floridi 2022.

<sup>22</sup> Floridi (cfr. *ivi*: 94-95) indica le seguenti come "le serie di principi più influenti per l'IA": i *Principi di Asilomar per l'IA* (Future of Life Institute, 2017); i principi contenuti nella *Dichiarazione di Montreal per l'IA responsabile* (University of Montréal, 2017); i principi contenuti nella seconda versione di *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Autonomous and Intelligent Systems* (IEEE, 2017); i principi contenuti nella *Dichiarazione su intelligenza artificiale, robotica e sistemi autonomi* (EGE, 2018) pubblicata dal Gruppo europeo sull'etica della scienza e delle nuove tecnologie della Commissione Europea; i "cinque principi generali per un codice di intelligenza artificiale" offerti nel rapporto *AI in the UK: Ready, Willing and Able?* (House of Lords – Artificial Intelligence Committee, 2017, § 417); i *Principi di partenariato dell'IA* (Partnership on AI, 2018).

<sup>23</sup> Cfr. Floridi 2022: 93-105. Per una trattazione più puntuale e approfondita, si veda Floridi et al. 2018.

<sup>24</sup> Floridi 2022: 102.

<sup>25</sup> *Ivi*: 96.

normativi, un “conflitto di doveri”<sup>26</sup>, o quantomeno di diritti. In questi casi, risulta evidente che non si tratta di stabilire a priori una regola di precedenza, né di riconoscere, all’esito di un bilanciamento, il principio di volta in volta prevalente. Quand’anche fosse, per giustificare una soluzione di questo tipo in modo convincente, occorrerebbe comunque disporre di una concezione unitaria e inclusiva dell’essere umano, dell’umanità nel suo complesso, dei suoi valori e scopi, che sia in grado di garantirne la superiorità e che a sua volta dimostri la propria validità, e preferibilità rispetto alle prospettive concorrenti, precisamente sottoponendosi al *test* circa la capacità di fornire risposte soddisfacenti alle problematiche connesse alla guida dell’IA, secondo una modalità tale per cui – per richiamare una nota formula<sup>27</sup> - la stessa umanità, in tutti i suoi membri e in tutte le sue fasi, forme e condizioni, possa essere trattata sempre anche come fine e mai semplicemente come mezzo.

## 6. La prospettiva personalista

Giunti a questo punto, potrebbe sembrare di trovarci ancora nelle nebbie, data la molteplicità di concezioni antropologiche esistenti e ugualmente adattabili al nostro scopo, parimenti sostenibili con validi argomenti e contestabili con obiezioni altrettanto valide. A ben vedere, tuttavia, qualche indicazione può essere ricavata dalla rassegna che ne abbiamo fatto.

Possiamo ritenere che, in una prospettiva orientata al recupero della centralità del soggetto umano nell’impiego delle tecnologie informatiche e dell’IA, non sembra poter essere adottata un’antropologia derivante da una visione utilitarista (in quanto essa determinerebbe la strumentalizzazione o l’asservimento di almeno alcuni esseri umani<sup>28</sup>), e neppure un’antropologia deontologista, imperniata sul rispetto di determinate regole (in quanto finirebbe per basarsi su un’assiologia particolare e risulterebbe quindi inevitabilmente oppressiva o discriminatoria nei confronti delle altre<sup>29</sup>) o giusnaturalista (perlomeno nella versione normativista, per l’astrattezza che il suo universalismo necessariamente comporterebbe<sup>30</sup>). Si potrebbe piuttosto ipotizzare che un’antropologia *personalista*<sup>31</sup>, sostenuta da una giustificazione filosofico-dialettica<sup>32</sup>, sia quella che consente di guardare all’essere umano non semplicemente come fine, ma come un fine che è

<sup>26</sup> Cfr. Grandi e Scantamburlo 2021: 50.

<sup>27</sup> Ci si riferisce evidentemente alla formula del secondo imperativo categorico kantiano: “Agisci in modo da trattare l’umanità, così nella tua persona come nella persona di ogni altro, sempre anche come fine, mai semplicemente come mezzo” (Kant 1997 [1785]: 91). Si noti come Kant, utilizzando l’espressione “mai semplicemente” (*niemals bloß*), manifesti la consapevolezza accorta e prudente che nelle relazioni sociali l’umanità possa essere trattata ‘anche’ come ‘mezzo’, ma nello stesso tempo escluda ‘categoricamente’ che possa essere trattata ‘semplicemente’ come tale. Circa la convergenza sul punto tra concezione kantiana e concezione aristotelica, si veda ancora Da Re 2010: specialm. 89-90.

<sup>28</sup> Cfr. Fabris 2021: 112.

<sup>29</sup> Cfr. Quintarelli (cur.) 2020: 98.

<sup>30</sup> Cfr. Grandi e Scantamburlo 2021: 50.

<sup>31</sup> Si noti che con questa espressione ci si riferisce ad un’antropologia della persona in generale e non al ‘personalismo’, inteso come corrente di pensiero sviluppatasi in Europa, soprattutto in Francia, nel secolo scorso, tra il primo e il secondo dopoguerra. Riguardo a quest’ultimo vanno richiamate le critiche di Berti, il quale opportunamente sottolinea come i filosofi ‘personalisti’ tendessero a concepire la persona come attività, negandole la sostanzialità, e così escludendo dal suo campo semantico quegli esseri umani che tale attività attualmente non svolgono, o non possono svolgere (cfr. Berti 1993b).

<sup>32</sup> Si veda, quale esemplificazione di tale giustificazione, Berti 1993a: 13-25.

fine a sé stesso, e che quindi è chiamato a realizzarsi finalisticamente<sup>33</sup>. Come ci ha insegnato la tradizione aristotelica, infatti, “la natura umana personale è teleologicamente orientata alla realizzazione di sé”<sup>34</sup> e non può quindi subordinarsi a sistemi tecnologici che in qualsiasi modo, per qualsiasi motivo, precludano, o riducano, questa possibilità. Verso un’antropologia di questo tipo, del resto, sembrano convergere le revisioni delle principali concezioni antropologiche che abbiamo pocanzi richiamato (§4).

### 7. L’insidia riduzionistica

Solo un’antropologia personalista risulta del resto in grado di fronteggiare la più minacciosa insidia del nostro tempo. Per quanto risalente perlomeno agli scritti di Lamettrie<sup>35</sup>, assistiamo infatti oggi al diffondersi ormai quasi onnipervasivo della tendenza a concepire l’essere umano come una macchina, sebbene ancora imperfetta. Come osserva Federico Faggin, uno dei grandi pionieri della civiltà informatica in cui viviamo,

“la maggior parte degli scienziati crede che siamo solo macchine: sofisticati sistemi di elaborazione delle informazioni basati su *wetware*. Ecco perché pensano che sarà possibile realizzare macchine che supereranno gli esseri umani. Credono che la coscienza emerga solo dal cervello, che sia prodotta da qualcosa di simile al software che funziona nei nostri computer. Pertanto, con un software più sofisticato, i nostri robot finiranno per essere consapevoli”<sup>36</sup>.

Ciò di cui parla Faggin, in effetti, è l’esito di una particolare linea di sviluppo della civiltà dell’informazione, precisamente quella che origina dal progetto cibernetico di Norbert Wiener. Come risulta dal suo testo fondativo del 1948, *Cybernetics, or Control and communication in the animal and the machine*<sup>37</sup>, il matematico statunitense riteneva che esseri viventi e macchine fossero accomunati da un’identica struttura informazionale, misurabile e controllabile dalla nuova scienza da lui scoperta. Due anni dopo, in *The human use of human being*, questa tesi verrà estesa alle organizzazioni sociali, in cui le persone si trovano inserite come parti dello stesso meccanismo. L’opera, tuttavia, si conclude con la seguente amara constatazione:

“Ho parlato di macchine, ma non soltanto di macchine che possiedono cervelli di ottone e muscoli di ferro. Allorché le persone umane sono organizzate nel sistema che le impiega non secondo le loro piene facoltà di esseri umani responsabili, ma come altrettanti ingranaggi, leve e connessioni, non ha molta importanza il fatto che la loro materia prima sia costituita da carne e sangue. Ciò che è usato come un elemento di una macchina, è un elemento nella macchina”<sup>38</sup>.

Wiener, dunque, è ben lungi dall’essere un entusiasta propugnatore della ‘civiltà delle macchine’, appare anzi perfettamente consapevole delle tragiche conseguenze che deriverebbero da

---

<sup>33</sup> Si noti qui la convergenza con la prospettiva già richiamata (vedi, *supra*, n. 27) dell’imperativo kantiano, pur basato su un impianto teorico, formalistico-trascendentale, completamente diverso.

<sup>34</sup> Palazzani 1996: 257.

<sup>35</sup> Cfr. Lamettrie 1955 [1748]. Circa le più remote origini cartesiane del paradigma meccanicista, si vedano i saggi raccolti in Bonicalzi (cur.) 2006.

<sup>36</sup> Faggin 2019: 317-318.

<sup>37</sup> Cfr. Wiener 2017 [1948].

<sup>38</sup> Wiener 2012 [1950]: 228-229.

un 'cattivo' "uso umano degli esseri umani". Eppure, è proprio dal suo progetto che ha preso forma un mondo in cui naturale e artificiale tendono sempre più a confondersi, tutto è ridotto a informazione, e il singolo uomo, svuotato della sua corporeità e interiorità, non può essere altro che "uno degli innumerevoli snodi comunicativi della rete globale"<sup>39</sup>. Ciò, del resto, è inevitabile. L'essere umano, concepito come una macchina, non può che subordinarsi alle macchine più potenti, che siano sistemi sociali o le IA più evolute, e questo indipendentemente dal fatto che tale subordinazione si realizzi consapevolmente, come nelle teorie del trans-umanesimo e del post-umanesimo<sup>40</sup>, o inconsapevolmente, come nelle visioni distopiche dei tecnotalitarismi<sup>41</sup>.

## 8. Una diversità irriducibile

Ma l'uomo è solo questo?

Scriva ancora Faggin che se crediamo di essere solo macchine classiche, "rischiamo di diventare sempre più disconnessi dai nostri sentimenti più profondi, dando sempre più validità ai simboli astratti in un mondo superficialmente razionale e meccanico e divorziato sempre di più dal significato e dai valori umani"<sup>42</sup>. Ridotti a sistemi informatici senza significato, insieme al significato perdiamo la capacità di agire coscientemente e liberamente<sup>43</sup>.

Eppure, sta proprio in questo agire cosciente e libero la differenza principale tra il nostro e il modo di essere tipico delle macchine. Esse ne sono prive. Come abbiamo visto<sup>44</sup>, se parliamo di 'autonomia' delle macchine lo possiamo fare solo in senso metaforico e dobbiamo quindi precisare accuratamente cosa intendiamo quando usiamo questa espressione a loro riguardo. Nel presente frangente ci può aiutare a comprenderlo meglio una notazione di Adriano Fabris:

"Potremmo dire che la macchina [...] è certamente in grado di *auto-regolare* i propri processi, ma non è capace di *auto-regolarli*. È questo il significato specifico, e il limite, della sua 'autodeterminazione'"<sup>45</sup>.

Ciò ha importanti implicazioni sotto il profilo giuridico. A differenza di quanto accade nei dispositivi artificiali, perfino nelle IA più evolute, l'essere umano ha la possibilità di aderire o meno alle regole che gli sono date, di metterle addirittura in questione, optando eventualmente per regole differenti. Ma soprattutto, anche quando segue una regola, ha la capacità di farlo consapevolmente e responsabilmente, cogliendone e accettandone il significato, assumendolo come motivazione della propria azione<sup>46</sup>.

Questa peculiarità del modo umano di stare di fronte al diritto richiede indubbiamente un'antropologia personalista per essere giustificata. Solo un'antropologia della persona che garantisca la sua irriducibilità all'operare della macchina, appare in grado di far sì che essa, anche

<sup>39</sup> Leonzi e Ciofalo 2017: 21.

<sup>40</sup> Si veda in proposito, per una panoramica, Ranisch and Sorgner (ed.), 2014.

<sup>41</sup> Cfr. sul punto Anders 2003: 58.

<sup>42</sup> Faggin 2019: 281.

<sup>43</sup> "Questo perché – osserva Faggin - l'informazione astratta senza significato che descrive il campo quantistico della fisica non può descrivere l'informazione inseparabile dal suo significato che esiste nei sé coscienti ed è controllata dal loro libero arbitrio" (ivi: 274).

<sup>44</sup> Cfr., *supra*, nota 7.

<sup>45</sup> Fabris 2018: 78.

<sup>46</sup> Insiste su queste implicazioni Fabris 2018: 120-121.



sotto il profilo giuridico, possa continuare ad essere trattata, in tutti i suoi membri e in tutte le sue fasi, forme e condizioni, come fine e non semplicemente come mezzo.

### 9. Quale personalismo?

Proprio a questo proposito si pongono tuttavia ulteriori questioni. Esistono infatti diverse specie di personalismo, ciascuna facente leva su un particolare aspetto della soggettività umana<sup>47</sup>. Di quale personalismo abbiamo dunque bisogno? Appare evidente che ne occorra uno che risulti quantomeno compatibile con la concezione dell'essere umano sopra (§6) delineata. Ma cosa significa questo in concreto? Quali caratteristiche sono necessarie per poter parlare di "persona" nel senso indicato? Un'indicazione importante al riguardo ci viene da Tommaso d'Aquino, il quale rielabora l'antica definizione boeziana, "*rationalis naturae individua substantia*"<sup>48</sup>, sottolineando la suprema eccellenza di un essere così concepito e motivandola con il suo carattere di soggetto *subsistens*, ovvero esistente in sé e per sé:

*"persona significat id quod est perfectissimum in tota natura, scilicet subsistens in rationali natura"*<sup>49</sup>; *"modus existendi quem importat persona est dignissimum, ut scilicet qliquid per se existens"*<sup>50</sup>.

Orbene, mentre l'essere *subsistens* indica l'essere in sé, non in altro, quindi l'essere sostanza, soggettività, non l'essere espressione o accidente di altro, viceversa l'essere *per se existens* indica l'essere per sé, l'essere in vista di sé, non in vista di altro, cioè l'essere fine, non mezzo.

La persona umana, dunque, benché dipendente da altro, essendo stata creata da Dio, è sussistente in sé ed esistente per sé, poiché dal Creatore è stata dotata di un essere ad essa proprio (*subsistens*) e di un fine ad essa intrinseco (*per se existens*)<sup>51</sup>.

Notiamo poi che la *perseità* della persona, il suo essere *per se existens*, trova compiuta possibilità di attuazione nella *razionalità* della sua *natura*, ovvero nella sua capacità di riflettersi come essere intelligente e libero, e quindi di autodeterminarsi in vista della propria realizzazione, del proprio compimento, della propria felicità, ovvero, per dirla con una parola, di trattarsi come "fine"<sup>52</sup>.

Dopo la crisi in cui è entrata in epoca moderna<sup>53</sup>, ci si può evidentemente chiedere se la concezione boeziano-tomista della persona sia riproponibile ai nostri giorni, così come si può discutere sull'accettabilità delle nozioni metafisiche implicate da questa concezione. Eppure, proprio dal dibattito contemporaneo<sup>54</sup> sembra emergere la necessità di ricorrere a tali nozioni se si vuole fornire adeguato fondamento ad una concezione dell'essere umano quale soggetto di diritti incompressibili, indisponibile a qualsiasi strumentalizzazione e manipolazione, anche tecnologica<sup>55</sup>.

<sup>47</sup> Si rimanda in proposito a quanto detto *supra* in §6, n. 31. Per ulteriori approfondimenti si veda De Anna 2007.

<sup>48</sup> Boezio 1979: 326 [*Contra Eutychem et Nestorium*: III, 1-6]

<sup>49</sup> Tommaso d'Aquino, *Summa Theologiae* I, q. 29, a. 3.

<sup>50</sup> Tommaso d'Aquino, *De Potentia*, q. 9, a. 4.

<sup>51</sup> Cfr. Berti 1993b: 32. Si veda anche, per una trattazione sistematica, Degl'Innocenti 1967.

<sup>52</sup> Cfr. Berti 1993b: 44-45.

<sup>53</sup> Cfr., per uno sguardo sintetico, De Anna 2007: 95 ss.

<sup>54</sup> Si veda per tutti, nell'ambito della filosofia anglo-americana di tradizione analitica, Strawson 2008: 95-121.

<sup>55</sup> Il punto è sviluppato in Berti 2004.

## 10. Conclusioni: il compito della filosofia

Emerge, in ogni caso, l'importanza di tornare a riflettere sulla natura dell'essere umano: se pensiamo, come si dice spesso<sup>56</sup>, all'IA come “uno specchio in cui guardiamo noi stessi”, allora “la bellezza o la mostruosità dell'immagine riflessa dipende forse interamente da noi”<sup>57</sup>. In fondo, potremmo anche vedere l'IA semplicemente come un'occasione e un banco di prova per una rinnovata ricerca filosofica su chi siamo.

Mai come ora, dunque, occorre capire chi è l'uomo e come deve essere trattato per salvaguardarne la dignità. Da qui la necessità di “rispettare l'ecologia umana»: tenere l'uomo «al centro della riflessione e come fine che qualifica il progresso” delle tecnologie informatiche, come scrivono Paolo Benanti e Sebastiano Maffettone<sup>58</sup>. Da qui anche la necessità di riattivare un'antropologia filosofica, e, più in generale, di riscoprire l'importanza della filosofia. Come scrive Bodei:

“il bisogno di comprendere i fini, oltre che i mezzi delle nuove tecnologie e delle loro ricadute sull'individuo, la società e la storia, sta inducendo le comunità scientifiche e le industrie ad accorgersi, come effetto collaterale, dell'utilità delle discipline umanistiche e, in particolare, della filosofia”<sup>59</sup>.

Soprattutto – potremmo aggiungere – della filosofia del diritto, che assume in questa prospettiva uno specifico rilievo, se non altro con riferimento alla teoria dei diritti fondamentali e alle condizioni della loro tutela.

Non basta semplicemente pensare a un umanesimo digitale, o a un'IA antropocentrica<sup>60</sup>. Già questa ormai non si può considerare più una prospettiva scontata, se solo si pensa alle visioni del transumanesimo e post-umanesimo. Ma anche rimanendo all'interno dell'orizzonte umanista, si tratta di interrogarsi su quale tipo o concezione di essere umano dovremmo adottare come modello di riferimento. Se sia ancora opportuno continuare a perseguire l'ideale utilitarista, inevitabilmente strumentalizzante almeno alcuni dei soggetti umani coinvolti, quello normativista, nelle sue varianti deontologista e naturalista, comunque insoddisfacenti – per parzialità o astrattezza - sul piano operativo, oppure se non sia necessario recuperare la prospettiva personalista, perlomeno in una versione che consenta all'essere umano di venire sempre trattato, in tutti i suoi membri e in tutte le sue fasi, forme e condizioni, anche come fine e mai semplicemente come mezzo.

<sup>56</sup> Si veda in proposito, per una disamina dei significati che può assumere questa immagine, Fabris 2021: 26 ss.

<sup>57</sup> Quintarelli (cur.) 2020: 100.

<sup>58</sup> Cfr., su questi temi, Benanti e Maffettone 2022: 29.

<sup>59</sup> Bodei 2019: 308-309.

<sup>60</sup> Sull'argomento esiste già un'ampia bibliografia. Segnalo in generale: Ford, Hayes, Glymour, and Allen 2015; Wei Xu 2019; Zhang, Bellamy, and Varshney 2020. In particolare, sui temi del *machine learning* e dell'*explainability*, si vedano: Riedl 2019; Springer 2019; Ehsan and Riedl, 2020. Può essere utile, infine, consultare il sito dello Stanford Institute for Human-Centered Artificial Intelligence (<https://hai.stanford.edu/>).

## BIBLIOGRAFIA

Ancona E. (cur.) 2019, “Soggettività, responsabilità, normatività 4.0. Profili filosofico-giuridici dell’intelligenza artificiale”, *Rivista di filosofia del diritto*, 8: 81-85.

Anders G. 2003 [1964], *Noi figli di Eichmann*, trad. di A.G. Saluzzi, Firenze: Giuntina.

Baron M.W., Pettit P., Slote M. 1997, *Three Methods of Ethics*, Oxford: Blackwell.

Benanti P., Maffettone S. 2022, “Decisioni politiche e sostenibilità digitale”, in P. Severino (cur.), *Intelligenza artificiale. Politica, economia, diritto, tecnologia*, Roma: Luiss University Press: 9-30.

Berti E. 1993a [1991], “L’etica alla ricerca della persona”, in E. Berti, *Soggetti di responsabilità. Questioni di filosofia pratica*, Reggio Emilia: Diabasis: 13-25.

Berti E. 1993b [1992], “Il concetto di persona nella storia del pensiero filosofico”, in E. Berti, *Soggetti di responsabilità. Questioni di filosofia pratica*, Reggio Emilia: Diabasis: 27-49.

Berti E. 2004, “Persona, scienza e tecnica”, in G. Galeazzi e B.M. Ventura (cur.), *Filosofia e scienza nella civiltà tecnologica*, Milano: F. Angeli: 171-183.

Bodei R. 2019, *Dominio e sottomissione. Schiavi, animali, macchine, Intelligenza Artificiale*, Bologna: Il Mulino.

Boezio, Severino M. 1979, *La consolazione della filosofia. Gli opuscoli teologici*, Milano: Rusconi.

Bonicalzi F. (cur.) 2006, *Macchine e vita nel XVII e XVIII secolo*, Firenze: Le Monnier.

Casadei T., Pietropaoli S. 2021, “Intelligenza artificiale: fine o confine del diritto?”, in T. Casadei, S. Pietropaoli (cur.), *Diritto e tecnologie informatiche. Questioni di informatica giuridica, prospettive istituzionali e sfide sociali*, Milano: Wolters Kluwer: 219-232.

Da Re A. 2010, *Le parole dell’etica*, Milano: Bruno Mondadori.

Danaher J. et al. 2017, “Algorithmic Governance: Developing a Research Agenda through the Power of Collective Intelligence”, *Big Data & Society*, 4 (2): 1-21.

De Anna G. 2007, “Persona: analisi storico-critica di una babele filosofica”, in G. Boniolo, G. De Anna, U. Vincenti, *Individuo e persona. Tre saggi su chi siamo*, Milano. Bompiani: 61-137.

Degl’Innocenti 1967, *Il problema della persona nel pensiero di S. Tommaso*, Roma: Libreria editrice della Pontificia Università Lateranense.

Ehsan U., Riedl M.O. 2020, "Human-centered Explainable AI: Towards a Reflective Sociotechnical Approach", *ArXiv*, 2002.01092: 1-19.

Eliot T.S. (ed.) 1961, *A Choice of Kipling's Verse*, London: Faber & Faber.

Fabris A. 1988, *Introduzione a M. Heidegger 1988 [1984]*, *Filosofia e cibernetica*, a cura di A. Fabris, Pisa: ETS.

Fabris A. 2018, *Etica per le tecnologie dell'informazione e della comunicazione*, Roma: Carocci.

Fabris A. 2021, *Etica delle nuove tecnologie*, Brescia: Scholé (Morcelliana).

Faggin F. 2019, *Silicio. Dall'invenzione del microprocessore alla nuova scienza della consapevolezza*, Mondadori, Milano.

Floridi L. 2022, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, ediz. it. a cura di M. Durante, Milano: Raffaello Cortina.

Floridi L. et al. 2018, "AI4People – An Ethical Framework for a Good AI Society. Opportunities, Risks, Principles, and Recommendations", *Minds and Machines*, 28 (4): 689-707.

Ford K.M., Hayes P.J., Glymour C., and Allen J. 2015, "Cognitive Orthoses: Toward Human-Centered AI", *AI Magazine*: 5-8.

Galimberti U. 1999, *Psiche e techne. L'uomo nell'età della tecnica*, Milano: Feltrinelli.

Grandi G., Scantamburlo T. 2021, "Apprendimento automatico e decisione umana", in F. Fossa, V. Schiaffonati, G. Tamburrini (cur.), *Automi e persone. Introduzione all'etica dell'intelligenza artificiale e della robotica*, Roma: Carocci: 29-53.

Head S. 2014, *Mindless. Why Smarter Machines Are Making Dumber Humans*, New York: Basic Books.

Hegel G.F.W. 1973 [1807], *Fenomenologia dello spirito*, trad. di E. De Negri, Firenze: La Nuova Italia.

Heidegger M. 1988 [1984], *Filosofia e cibernetica*, a cura di A. Fabris, Pisa: ETS,.

Kant I. 1997 [1785], *Fondazione della metafisica dei costumi*, trad. e introd. di F. Gonnelli, Bari-Roma: Laterza.

Knight W. 2017, "The Dark Secret at the Hearth of IA", *MIT Technology Review (April 11)*. Available at: <https://www.Technology-review.com/s/604087/the-dark-secret-at-the-heart-of-ai/>.

Kurzweil R. 2005, *The Singularity is Near: When Humans Transcend Biology*, New York: Viking press.

Lamettrie J.O. de 1955 [1748], “L’uomo macchina”, in J.O. de Lamettrie, *L’uomo macchina ed altri scritti*, a cura di G. Preti, Milano: Feltrinelli.

Leonzi S., Ciofalo G. 2017, “Wiener, Goldrake e i Social Media. L’eredità della cibernetica tra scienza, tecnologia e comunicazione”, in N. Wiener, *La cibernetica*, trad. e cur. di S. Leonzi e G. Ciofalo, Roma: Armando: 7-31.

P8\_TA (2017) 0051. *Norme di diritto civile sulla robotica. Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica* (2015/2103 [INL]).

Palazzani L. 1996, *Il concetto di persona tra bioetica e diritto*, Torino: Giappichelli.

Palazzani L. 2020, *Tecnologie dell’informazione e intelligenza artificiale. Sfide etiche al diritto*, Roma: Studium.

Pirosa R. 2021, “I dilemmi della tecnoregolazione: uno sguardo critico”, in T. Casadei, S. Pietropaoli (cur.), *Diritto e tecnologie informatiche. Questioni di informatica giuridica, prospettive istituzionali e sfide sociali*, Milano: Wolters Kluwer: 275-286.

Quintarelli S. (cur.) 2020, *Intelligenza artificiale. Cos’è davvero, come funziona, che effetti avrà*, Torino: Bollati Boringhieri.

Ranisch R., Sorgner S.L. (eds) 2014, *Post- and Transhumanism: An Introduction*, Frankfurt a.M.: Peter Lang.

Riedl M.O. 2019, “Human-centered artificial intelligence and machine learning”, *Human Behavior and Emerging Technologies*: 33– 36.

Severino E. 2008, *La tendenza fondamentale del nostro tempo*, in E. Severino, *La tendenza fondamentale del nostro tempo*, Milano: Adelphi: 37-66.

Simoncini A. 2019, “L’algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà”, *Biolaw Journal – Rivista di BioDiritto*, 1: 63-89.

Singer P. (ed.) 2006, *A Companion to Ethics*, Malden (MA): Wiley-Blackwell.

Springer A. 2019, *Accurate, Fair, and Explainable: Building Human-Centered AI*. PhD Thesis, Santa Cruz: University of California.

Strawson P.F. 2008 [1959], *Individui. Saggio di una metafisica descrittiva*, trad. di E. Bencivenga, Milano-Udine: Mimesis.

Tamburrini G. 2020, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Roma: Carocci.

Vinge V. 1993, “The Coming Technological Singularity: How to Survive in the Post-Human Era”, in *Vision-21. Interdisciplinary Science and Engineering in the Era of Cyberspace. Proceedings of a symposium cosponsored by the NASA Lewis Research Center and the Ohio Aerospace Institute and held in Westlake (Ohio): 11-22*. Available at: <http://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html>.

Wei Xu 2019, “Toward Human-centered AI: a Perspective from Human-Computer Interaction”, *Interactions*, 26 (4): 42–46.

Wiener N. 2012 [1950], *Introduzione alla cibernetica. L'uso umano degli esseri umani*, trad. di D. Persiani, Torino: Bollati Boringhieri.

Wiener N. 2017 [1948], *La cibernetica*, trad. e cur. di S. Leonzi e G. Ciofalo, Roma: Armando.

Zhang Y., Bellamy R.K.E., Varshney K.R. 2020, “Joint Optimization of AI Fairness and Utility: A Human-Centered Approach”, in *AIES '20: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, New York (NY): Association for Computing Machinery: 400–406.