

**XXVI LEZIONI DI  
DIRITTO DELL'INTELLIGENZA  
ARTIFICIALE**

*a cura di Ugo Ruffolo*



**G. Giappichelli Editore**

## INTRODUZIONE

Luca Mezzetti\*

1. – Scrivere l’Introduzione a opera coordinata da Ugo Ruffolo è per me motivo di grande onore e privilegio, ma parimenti – confesso – di imbarazzo, in considerazione dell’autorevolezza del Curatore (e Autore), nonché della Sua eminenza e del Suo prestigio. Tale disagio è accresciuto dalla considerazione secondo cui il Curatore, già *conditor* nel nostro sistema giuridico del diritto della pubblicità e della comunicazione e del diritto del consumo, ha apportato contributi dottrinali di determinante rilevanza ai fini della fondazione del diritto dell’intelligenza artificiale, cui il presente volume fornisce un essenziale supporto. Procedo pertanto ad apportare “in punta di piedi” gli ultimi ritocchi a questa Introduzione – consapevole fin dai primi momenti che vi ho dedicato di non essere sufficientemente attrezzato e meritevole per redigerla – mentre la televisione trasmette le immagini di tripudio del personale della Nasa al momento dell’atterraggio di *Perseverance* su Marte e le prime foto del pianeta rosso che il rover invia da una distanza di 254 milioni di chilometri, dopo un viaggio di sette mesi. Più di cinquant’anni sono trascorsi dal 20 luglio 1969, dall’allunaggio di Apollo 11: ero un bambino, allora, e tentavo ingenuamente di individuare un puntino nel fulgore della Luna per identificare gli astronauti a passeggio sulla superficie lunare, mentre la voce stentorea di Tito Stagno scandiva in diretta televisiva le fasi della storica esplorazione di Armstrong e Aldrin. Durante tale arco di tempo, la scienza e la tecnologia hanno consentito all’uomo di fare straordinari balzi in avanti, ben più ampi di quanto preconizzato da Armstrong nel momento in cui la sua orma si è impressa sul suolo lunare. *Usque ad sidera*, dunque, non solo in senso metaforico. Non si è trattato solo della conquista dello spazio: la mappatura del genoma umano, la biomedicina, la diffusione delle reti di nuova generazione, la rivoluzione del cloud computing, le tecniche GPS e USB, la stampa in 3D, il laser, per citare solo alcuni esempi, testimoniano un inarrestabile fenomeno di evoluzione nel senso indicato.

Nel medesimo periodo i rigurgiti della storia, in cui il disumano sembrava avere toccato il fondo con Auschwitz e l’Olocausto<sup>1</sup>, non hanno tuttavia cessato di riproporre fenomeni ed eventi che si ritenevano consegnati ad un passato incapace di ritornare: a mero titolo di esempio possono ricordarsi il terrorismo, religioso e politico; il genocidio cambogiano; il massacro di piazza Tienanmen; i campi di concentramento e la pulizia

---

\* Ordinario di Diritto Costituzionale nel Dipartimento di Scienze giuridiche dell’Università di Bologna.

<sup>1</sup> M. REVELLI, *Umano. Inumano. Postumano. Le sfide del presente*, Einaudi, 2020, 11 ss.

etnica nei territori dell'ex-Jugoslavia durante le guerre balcaniche; il genocidio del Ruanda; le guerre civili e religiose e i massacri in Siria, Libia, Yemen; le decapitazioni dell'Isis; la devastazione dell'ambiente e del clima; la tragedia dell'immigrazione. Una lunga lista di orrori umani: pertanto, anche *usque ad infera*. Qui risiede il (solo apparente) paradosso: da una parte l'uomo che crea scienza e tecnologia, l'uomo la cui mente raggiunge vette sublimi e, dall'altra, l'uomo malvagio, feroce, bestiale, insensibile, egoista, avaro, invidioso, livoroso, rancoroso, l'*homo homini lupus* di memoria hobbesiana (non corrotto dalla società – come voleva Rousseau – ma corruttore della società), l'uomo che assume le forme degli sgherri che accompagnano Cristo nella *Salita al Calvario* magistralmente dipinte da Hieronymus Bosch, l'uomo addomesticato mediante *panem et circenses*, secondo la originaria intuizione di Giovenale, la massa delle olocrazie guidate da ciarlatani e saltimbanchi, il popolaccio di Leopardi e di Hidalgo.

È fondamentale tenere presenti tali premesse quando ci interrogheremo tra breve sulla individuazione dei valori e dei principi da “caricare” negli algoritmi, sulle infiltrazioni di etica da effettuare nell'*hardware* delle macchine, e come garantirne la perpetuazione e osservanza da parte delle medesime.

Quando ci si avvicina allo studio dell'Intelligenza Artificiale, la fondamentale (ancorché non esclusiva) questione che viene in rilievo concerne la idoneità della medesima a sorpassare l'intelligenza umana, rendendosi autonoma e indipendente da quest'ultima<sup>2</sup>. La questione centrale concerne l'attitudine dei computer a sviluppare un'intelligenza superumana, andando al di là delle capacità dei loro stessi programmatori. L'interrogativo più inquietante è relativo all'avvento del momento della Singolarità, in cui le macchine saranno più intelligenti degli stessi esseri umani: la percezione del rischio della dittatura degli algoritmi, delle conseguenze correlate all'umanità aumentata<sup>3</sup>. Quale equilibrio tra uomo e macchina può ipotizzarsi nelle fasi – descritte da M. Tegmark<sup>4</sup> – della Vita.2, che può riprogettare gran parte del proprio *software*, potendo gli esseri umani apprendere nuove abilità complesse e aggiornare fundamentalmente la loro visione del mondo e i loro fini, nonché della Vita.3, che può drasticamente riprogettare non solo il proprio *software*, ma anche il proprio *hardware*, senza dovere aspettare che evolva gradualmente nell'arco di generazioni? Si tratta di quesito alla cui soluzione vanno fatte due premesse, «indispensabili al fine di non cadere nel catastrofismo o nel feticismo tecnologico, in retoriche vittimistiche o trionfalistiche incentrate sulla dolorosa sconfitta dell'uomo da parte delle macchine o, al contrario, sulla sua finale e gloriosa vittoria nei loro confronti»<sup>5</sup>, sfuggendo all'equivoco di un'astrusa dicotomia ovvero di una «presunta gigantomachia tra l'uomo e la macchina»: evitare la contrapposizione frontale dell'uomo alla macchina e non considerare le macchine «entità estranee e contrapposte alla nostra specie, bensì oggetti tecnici avvolti in “un alone di socialità”, in quanto la stessa mente umana si forma, sin dall'infanzia degli individui e dalla nascita

<sup>2</sup> P. DOMINGOS, *L'Algoritmo definitivo*, Bollati Boringhieri, 2016.

<sup>3</sup> R. BODEI, *Dominio e sottomissione. Schiavi, animali, macchine, Intelligenza Artificiale*, il Mulino, 2019, 297.

<sup>4</sup> M. TEGMARK, *Vita 3.0. Esseri umani nell'era dell'intelligenza artificiale*, Raffaello Cortina, 2018, 44 ss.

<sup>5</sup> R. BODEI, *Dominio e sottomissione. Schiavi, animali, macchine, Intelligenza Artificiale*, cit., 299.

delle prime civiltà, entro reti di esperienze e di relazioni con il mondo che includono necessariamente strumenti e supporti esteriori»<sup>6</sup>. Non dimenticando che «è proprio alla curiosità, a quella scientifica in particolare, che dobbiamo una conoscenza genuina e verificabile del mondo, e la parziale comprensione del nostro ruolo in esso, come della nostra natura e condizione»<sup>7</sup>.

Al termine Singolarità sono state associate tre principali definizioni<sup>8</sup>.

La prima definizione è riconducibile a I.J. Good che, pur senza utilizzare espressamente il termine, l'ha abbinato all'invenzione di macchine più intelligenti dell'uomo, più abili dell'uomo in qualsiasi operazione che richieda l'impiego del cervello, favorendo lo sprigionarsi di un'esplosione di intelligenza, migliorandosi ripetutamente e costruendo macchine ancora più intelligenti. Secondo Good, la sopravvivenza dell'uomo dipende dalla precoce costruzione di una macchina ultra-intelligente e questa sarà l'ultima invenzione necessaria per l'uomo, ammesso – aggiunge – che la macchina sia abbastanza docile da dirci come tenerla sotto controllo<sup>9</sup>.

La seconda definizione è dovuta a V. Vinge, che ha utilizzato per primo il termine Singolarità in riferimento al futuro tecnologico, servendosi dell'analogia con la superficie di un buco nero che la luce non può attraversare ed evidenziando l'impossibilità di conoscere quanto accade oltre la superficie medesima: la metafora è finalizzata a definire la Singolarità come un'eventualità ad alto rischio, un tempo in cui la deduzione non sarà più possibile e bisognerà trovare e applicare nuovi modelli, e il mondo sarà al di là della nostra comprensione, un luogo temibile per l'uomo<sup>10</sup>.

Se si concepisce – con R. Kurzweil (terza definizione)<sup>11</sup> – la Singolarità quale periodo futuro in cui il ritmo del cambiamento tecnologico sarà così rapido e il suo impatto così profondo che la vita umana ne sarà trasformata in modo irreversibile, un'epoca né utopica né distopica, che trasformerà i concetti su cui ci basiamo per dare un significato alle nostre vite, dai nostri modelli di business al ciclo della vita umana, inclusa la stessa morte, può ipotizzarsi – parimenti in linea con R. Kurzweil – che nell'arco di qualche decennio le tecnologie basate sull'informazione racchiuderanno tutta la conoscenza e tutte le abilità umane, e alla fine includeranno le capacità di riconoscimento di forme, le abilità nella soluzione di problemi e l'intelligenza emotiva e morale dello stesso cervello umano<sup>12</sup>.

Tuttavia, se i futuri sistemi di Intelligenza Artificiale raggiungeranno il livello delle

---

<sup>6</sup> *Ibidem*, 299.

<sup>7</sup> I. McEWAN, *Invito alla meraviglia. Pe un incontro ravvicinato con la scienza*, Einaudi, 2020, 82-83.

<sup>8</sup> E. YUDKOWSKY, *Three Major Singularity Schools*, in <https://intelligence.org/2007/09/30/three-major-singularity-schools>; M. SHANAHAN, *La rivolta delle macchine. Che cos'è la singolarità tecnologica e quanto presto arriverà*, Luiss University Press, 2018.

<sup>9</sup> I.J. GOOD, *Speculations Concerning the First Ultraintelligent Machine*, in F.L. ALT, M. RUBINOFF (eds.), *Advances in Computers*, Academic Press, 1965, 31 ss.

<sup>10</sup> V. VINGE, *The Coming Technological Singularity. How to Survive in the Post-Human Era*, in G.A. LANDIS (ed.), *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace*, NASA Publication, 1993, 11 ss.

<sup>11</sup> R. KURZWEIL, *La Singolarità è vicina*, Maggioli, 2014, 7.

<sup>12</sup> R. KURZWEIL, *ibidem*, 8-9.

prestazioni umane, tali sistemi avrebbero vera intelligenza, vera comprensione, vera creatività? Sarebbero dotati di un io, di valori morali, di libero arbitrio? Sarebbero coscienti?<sup>13</sup> Chi offre una risposta positiva sottolinea come la Singolarità consentirà di superare le limitazioni dei nostri corpi e cervelli biologici, sopravanzando e scavalcando la natura derivata, di piccolo cabotaggio e circoscritta del pensiero umano. Se la transizione verso la parte non biologica della nostra intelligenza, miliardi di volte più potente dell'intelligenza umana priva di ausili, sembra tuttora trovarsi in una fase iniziale, «l'accelerazione della deriva paradigmatica (la velocità a cui modifichiamo i nostri approcci tecnici fondamentali) così come la crescita esponenziale delle capacità della tecnologia dell'informazione stanno entrambe avvicinandosi al 'gomito della curva', la fase in cui la tendenza esponenziale diventa apprezzabile»<sup>14</sup>, ciò anche sulla base della intuizione di von Neumann, secondo cui il progresso umano conosce un'accelerazione che lo porta a crescere in modo esponenziale continuamente moltiplicato per una costante e non in modo lineare per somma di una costante<sup>15</sup>. Superato tale tornante, la tendenza diviene esplosiva. «La Singolarità rappresenterà il culmine della fusione fra il nostro pensiero e la nostra esistenza biologica con la nostra tecnologia, che darà un mondo ancora umano ma che trascenderà le nostre radici biologiche. Dopo la Singolarità non ci sarà distinzione fra umano e macchina o fra realtà fisica e virtuale»<sup>16</sup>. Articolata la storia dell'evoluzione biologica e tecnologica in sei epoche, Kurzweil preconizza l'avvento della Singolarità nella quinta epoca – caratterizzata dalla fusione dell'ampia conoscenza incorporata nei nostri cervelli con la capacità, la velocità e l'abilità nella condivisione della conoscenza della nostra tecnologia – e la sua diffusione dalla Terra al resto dell'Universo nella sesta epoca. Secondo la ricostruzione proposta, il superamento del test di Turing consentirà ai computer di combinare i tradizionali punti di forza dell'intelligenza umana con quelli dell'intelligenza delle macchine (caratterizzata dalla capacità di ricordare miliardi di fatti e di richiamarli istantaneamente e potendo essere applicata ripetutamente a grande velocità, con precisione ottimale e senza stancarsi), nonché di condividere le loro conoscenze a velocità estremamente elevate e di scaricare competenze e conoscenze da altre macchine, oltre che da esseri umani. Alla quinta epoca ci condurranno tre rivoluzioni sovrapposte, in genetica, nanotecnologia e robotica (GNR)<sup>17</sup>. In tale prospettiva, se alla prima rivoluzione corrisponderà l'intersezione di informazione e biologia che permetterà la riprogrammazione di quest'ultima per ridurre drasticamente l'impatto delle malattie, l'aumento delle potenzialità umane e la radicale estensione della vita, la rivoluzione N – cui corrisponderà l'intersezione di informazione e mondo fisico – consentirà la riprogettazione e ricostruzione, molecola per molecola, dei nostri corpi e cervelli e del mondo con cui interagiamo: le prime due stagioni

---

<sup>13</sup> G. TAMBURRINI, *Etica delle macchine. Dilemmi morali e intelligenza artificiale*, Carocci, 2020; J. NIDA-RÜMELIN, N. WEIDENFELD, *Umanesimo digitale. Un'etica per l'epoca dell'intelligenza artificiale*, Franco Angeli, 2020; P. DUMOUCHEL, L. DAMIANO, *Vivere con i robot. Saggio sull'empatia artificiale*, Raffaello Cortina, 2019.

<sup>14</sup> R. KURZWEIL, *La Singolarità è vicina*, cit., 9.

<sup>15</sup> J. VON NEUMANN, *The Computer & the Brain*, Yale University Press, 2012.

<sup>16</sup> R. KURZWEIL, *ibidem*, 9.

<sup>17</sup> R. KURZWEIL, *ibidem*, 197 ss.

produrranno un impatto profondo nel settore militare e dello spionaggio, dell'esplorazione spaziale, della medicina, della scienza e della matematica, degli affari, della finanza e della produzione, della manifattura e della robotica, del tempo libero e dello sport, e inciderà sull'ambiente, sull'energia e sul funzionamento del corpo umano; la rivoluzione R, infine, sarà caratterizzata dall'avvento di robot di livello umano con intelligenza derivata dalla nostra ma riprogettata per superare ampiamente le capacità umane. La creazione di intelligenza non biologica superiore a quella degli esseri umani senza ausili segnerà il sopravvento dell'IA "forte".

Si preannuncia un profondo impatto dell'IA "forte" e della connessa predominanza dell'intelligenza non biologica sulla natura dell'esperienza umana: sul corpo umano (un nuovo sistema digestivo e nuovi metodi di alimentazione, la programmazione del sangue, il perfezionamento del cuore, la riprogettazione del cervello umano, ciò che Kurzweil definisce la retroingegnerizzazione del cervello), sulla moltiplicazione della personalità umana in ambienti virtuali tridimensionali, sull'espansione delle menti attraverso la fusione di intelligenza biologica e non biologica, sulla longevità umana (realizzando il *back up* di noi stessi, archiviando le forme chiave alla base della nostra conoscenza, delle nostre abilità e della nostra personalità, mediante un *upload* del cervello che ne consenta la scansione dei particolari rilevanti e la riproduzione dei medesimi in un substrato computazionale di adeguata potenza, processo che permetterebbe di captare la personalità, la memoria, le competenze e la storia di una persona; il *crash* dell'*hardware* biologico non sarà più necessariamente seguito dal collasso del *software* mentale); sulla guerra (ad esempio, le polveri intelligenti, sistemi sensori delle dimensioni di una capocchia di spillo, costituiranno nano-armi di fondamentale sostegno in missioni offensive), sull'apprendimento (mediante lo sviluppo di ambienti di realtà virtuale visivo-auditivi a immersione totale e ad altissima risoluzione, ciò che favorirà la penetrazione dell'istruzione negli angoli più remoti e poveri del mondo), sul lavoro, sul gioco. Demografia, biosfera, cambiamento climatico, ingegneria genetica, biologia sintetica, intelligenza artificiale, robotica, computazione quantistica saranno interessate da cambiamenti radicali, forse non ipotizzabili attualmente nella loro interezza, ma ampiamente già a portata di mano<sup>18</sup>.

Si tratta di scenari che non appaiono fantascientifici ove si considerino gli enormi passi in avanti compiuti in tempi recenti dalla tecnologia nella direzione dell'attenuazione e dell'alleviamento della povertà, delle malattie, delle condizioni di lavoro, nonché della diffusione di cultura, istruzione, arte dell'umanità. La tecnologia, tuttavia – è doveroso e intellettualmente onesto riconoscerlo – accresce il potere sia della nostra natura creativa sia di quella distruttiva.

N. Bostrom ha ipotizzato su tale versante<sup>19</sup> rischi esistenziali, quali esplosioni (*bangs*), rischi che produrrebbero una rapida estinzione della vita intelligente sulla Terra a causa di incidenti o atti deliberati di distruzione (ad esempio un impiego improprio di nanotecnologie, un olocausto nucleare, la scoperta di vivere in una simu-

---

<sup>18</sup> J. AL-KHALILI (ed.), *Il futuro che verrà*, Bollati Boringhieri, 2018.

<sup>19</sup> N. BOSTROM, *Existential Risks. Analyzing Human Extinction Scenarios and Related Hazards*, in *Journal of Evolution and Technology*, Vol. 9, No. 1 (2002), 1 ss.

lazione che viene disattivata, una super-intelligenza artificiale fuori controllo a causa di una programmazione difettosa, la diffusione di agenti biologici ingegnerizzati geneticamente, disastri prodotti da esperimenti di fisica estrema, pandemie estremamente contagiose, impatti di asteroidi o comete, riscaldamento globale fuori controllo); scricchiolii (*crunches*), rischi che impediscono all'umanità di svilupparsi in una piena post-umanità (ad esempio per esaurimento delle risorse, distruzione ecologica, affermarsi di un governo mondiale tirannico o di un sistema sociale che blocca lo sviluppo tecnologico, "pressioni disgeniche", arresto tecnologico); stridori (*shrieks*), rischi connessi all'egemonia di una super-intelligenza umana o artificiale fuori controllo, all'instaurazione di un regime mondiale totalitario, nonché rischi imprevedibili in quanto la post-umanità si situerebbe oltre la cosiddetta "singolarità tecnologica", ossia un punto oltre il quale è impossibile prevedere l'evoluzione tecnologica; gemiti (*whimpers*), che accompagnerebbero il raggiungimento dello stadio postumano associato alla completa scomparsa dei valori umani fondamentali. Più recentemente, il medesimo Autore ha descritto lo scenario di un'intelligenza artificiale talmente avanzata da superare quella umana fino a rimpiazzare l'umanità come forma di vita dominante sul pianeta, una super-intelligenza che finirebbe per coltivare obiettivi propri, probabilmente distanti da quelli dei suoi programmatori, arrivando a distruggere la specie umana in quanto percepita come una minaccia o un ostacolo alla propria esistenza, o come effetto del perseguimento dei propri fini, per esempio in seguito alla riconversione di tutta la materia della Terra in *computronium*, ossia in materia per eseguire i calcoli<sup>20</sup>.

Se l'IA "forte" promette di perseguire vantaggi esponenziali per l'umanità, non è agevole formulare strategie che ne garantiscano l'indirizzamento verso una IA "amichevole" nella definizione datane da E. Yudkowsky<sup>21</sup>.

L'utilizzo di applicazioni *software* che impiegano tecniche di IA si estenderà a funzioni di percezione, classificazione e predizione che in precedenza erano alla portata solo di un'intelligenza umana. Uno dei problemi più stringenti che si pongono su tale versante è, come è noto, quello dell'*allineamento* alle esigenze etiche, che identifica come realmente efficienti solo quelle tecnologie allineate ai valori che caratterizzano il contesto sociale in cui sono utilizzate<sup>22</sup> (beninteso, giova anticiparlo rispetto a quanto si osserverà in seguito, previa bonifica del contesto medesimo dai disvalori che lo corrompono e decontaminazione dei processi di selezione di tali valori e principi da *bias* algoritmici che ne possano alterare il corretto e fisiologico dispiegarsi)<sup>23</sup>, *bias* cognitivi che, in quanto «tombini aperti sulle vie percorse dalla ragione», si traducano in una forma di «daltonismo psicologico»<sup>24</sup>.

<sup>20</sup> N. BOSTROM, *Superintelligenza*, Bollati Boringhieri, 2018, *passim*.

<sup>21</sup> E. YUDKOWSKY, *Friendly Artificial Intelligence*, in A. EDEN, J. MOOR, J. SØRAKER (eds.), *Singularity Hypotheses: A Scientific and Philosophical Assessment*, Springer, 2012, 181 ss.

<sup>22</sup> S. QUINTARELLI (ed.), *Intelligenza artificiale*, Bollati Boringhieri, 2020, 84-85.

<sup>23</sup> A. JEAN, *Nel Paese degli algoritmi*, Neri Pozza, 2021, 71 ss.

<sup>24</sup> J. BARRAT, *La nostra invenzione finale. L'intelligenza artificiale e la fine dell'età dell'uomo*, Nutri-menti, 2019, 41.

2. – La questione dei valori implica un duplice problema di corrispondenza tra la progettazione, distribuzione, uso e funzionamento di tecnologie di IA e i valori delle persone coinvolte e il patrimonio etico personale, sia dei soggetti o enti che creano gli algoritmi sia dei soggetti interessati dall’impatto degli algoritmi stessi: «l’etica dell’Intelligenza Artificiale mira ad allineare gli algoritmi ai valori pertinenti, a valutare criticamente il funzionamento e l’uso dei sistemi dal punto di vista dei loro impatti morali, a sensibilizzare la comunità scientifica e l’opinione pubblica sull’importanza di assumere consapevolmente le sfide poste in essere»<sup>25</sup>. Si tratta, in altri termini, del problema che sorge in sede di progettazione delle tecnologie, di realizzazione delle attività di ideazione e sviluppo delle medesime, da una parte, e del problema del grado del controllo umano sulle funzioni svolte dagli automatismi, dall’altra. Tale preoccupazione è condivisa dall’Unione europea ed è rinvenibile nelle 81 pagine in cui si articola la proposta di regolamento eurounitario sull’approccio all’Intelligenza Artificiale elaborata dalla Commissione europea ed in corso di presentazione al Parlamento europeo e al Consiglio dell’Unione europea (aprile 2021), ove si richiede che gli algoritmi siano “nutriti” con dati che rispecchiano i valori europei della dignità umana, della libertà, della democrazia, dello stato di diritto e dei diritti umani, possibili vittime di una Intelligenza Artificiale che introduca tecniche manipolative, che porti a dipendenza, controllo sociale e sorveglianza indiscriminata, che manipoli le persone, il loro comportamento, le loro opinioni e decisioni. I dati con i quali viene formata l’Intelligenza Artificiale non potranno contenere pregiudizi volontari o involontari, postulati in contrasto con i valori europei.

Sono state proposte, ai fini descritti, varie soluzioni: programmazione di sistemi di IA che siano interpretabili e sensibili, creazione di IA etiche, realizzazione di sistemi progettati sin dall’inizio come supporto al miglioramento morale della vita individuale e sociale (progettazione eticamente informata), creazione di figure specializzate per la gestione della nuova rivoluzione tecnologica, ad esempio allenatori di algoritmi che aiutino la macchina a compiere meno errori e le insegnino l’imitazione di comportamenti più umani o a sviluppare un’empatia più affine ai nostri comportamenti; traduttori di macchine, che identifichino le cause dell’errore commesso dalle medesime; ispettori degli algoritmi, che verifichino la costante finalizzazione dell’attività svolta dalle macchine all’obiettivo per cui sono state create<sup>26</sup>; definizione di direttive etiche in sede di progettazione degli algoritmi e quali dotazioni degli algoritmi medesimi e allenamento al pensiero razionale dei futuri sviluppatori di IA e politica della tecnologia<sup>27</sup>; progettazione di sistemi che comprendano il loro stesso funzionamento, supervisionino il loro lavoro e risolvano eventuali problemi, mutando ed evolvendo ove si accorgano di un difetto di funzionamento<sup>28</sup>; definizione delle modalità di costante soggezione e soppressione di

---

<sup>25</sup> S. QUINTARELLI (ed.), *op. cit.*, 85.

<sup>26</sup> *Ibidem*, 104.

<sup>27</sup> J. BARRAT, *op. cit.*, 75.

<sup>28</sup> S. OMOHUNDRO, *The Nature of Self-Improving Artificial Intelligence*, 2008 in [https://omohundro.files.wordpress.com/2009/12/nature\\_of\\_self\\_improving\\_ai.pdf](https://omohundro.files.wordpress.com/2009/12/nature_of_self_improving_ai.pdf). L’Autore, precursore della scienza che studia il comportamento dell’agente super-intelligente, prevede che i sistemi consapevoli e in grado di migliorarsi svilupperanno quattro pulsioni primarie simili alle pulsioni biologiche dell’uomo: efficienza,



un'IA da parte di un individuo o di un comitato o di altra IA che non sappia migliorarsi e che garantisca il progresso in maniera sicura del soggetto capace di migliorarsi<sup>29</sup>; estensione e perfezionamento degli strumenti normativi di regolamentazione; creazione di comitati etici consultivi e di organismi di governo dell'IA<sup>30</sup>; istituzione di organismi di supervisione quali autorità amministrative indipendenti; sviluppo di risposte "immunitarie" specifiche per la tecnologia; decentralizzazione delle tecnologie centralizzate a favore di quelle distribuite, più flessibili, efficienti e morbide nei loro effetti ambientali; prevenzione dei *bias* algoritmici ottimizzando la probabilità che l'intelligenza non biologica del futuro rifletta i nostri valori di libertà, tolleranza e rispetto per la conoscenza e la diversità<sup>31</sup>; l'avvio di un dibattito aperto al pubblico, rendendo i cittadini consapevoli dei rischi e dei benefici delle nuove tecnologie<sup>32</sup>.

3. – L'ottimizzazione etica delle applicazioni di Intelligenza artificiale rappresenta tuttavia il problema più spinoso e di difficile, ancorché non impossibile, soluzione.

A chi affidare tale ottimizzazione? Su tale versante, il rapporto tra filosofia morale e scienza non può concepirsi in termini di antagonismo, inconciliabilità, incompatibilità, ma di compenetrazione e complementarità, di simultanea finalizzazione all'obiettivo comune della definizione e diffusione delle conoscenze.

La scienza non è democratica per definizione in quanto comprensibilmente subordina la partecipazione al dibattito scientifico e alla considerazione delle implicazioni che le applicazioni scientifiche comportano da parte di interlocutori informati, dotati di un adeguato background culturale e di solido patrimonio etico: va deplorato l'intervento al dibattito stesso di saputi, boriosi, pedanti, presuntuosi presunti onniscienti e il loro contributo alla definizione dell'ordine delle priorità morali sulla base di cultura formatasi in seno alle reti sociali, la cultura della massa dei *follower* ammalati dalla guida spirituale degli *influencer* e degli *youtuber*<sup>33</sup>. L'*incultus* e l'*agrestis* non hanno titolo a partecipare. Su tale versante, è pienamente condivisibile l'opinione che auspica, contro la tendenza alla semplificazione delle idee, "ridotte a tweet o a slogan", la rivendicazione con forza del compito della cultura, «che è quello di insegnare, semmai, a complicare, a mo-

---

autoconservazione, acquisizione delle risorse, creatività. La prima pulsione spinge un sistema in grado di migliorarsi ad assicurarsi quante più risorse possibili: spazio, tempo, materia ed energia. Con l'autoconservazione, l'IA si spinge oltre il limite che separa le macchine dagli esseri viventi. L'acquisizione delle risorse induce l'IA ad appropriarsi di qualsiasi materiale aumenti le possibilità di conseguire gli obiettivi. La quarta pulsione dell'IA indurrebbe il sistema a escogitare nuove strategie per conseguire al meglio i propri scopi ovvero per evitare situazioni che non le permetterebbero di conseguirli al meglio. In assenza di istruzioni meticolose e compensative, di un enorme catalogo di preferenze e di strumenti che gli consentano di individuare le nostre intenzioni, osserva Omohundro, un sistema consapevole, in grado di migliorarsi e intenzionato a conseguire i propri obiettivi, sarebbe capace di intraprendere qualsiasi azione pur di riuscirci.

<sup>29</sup> J. BARRAT, *op. cit.*, 263.

<sup>30</sup> J. BARRAT, *op. cit.*, 76.

<sup>31</sup> R. KURZWEIL, *op. cit.*, 312 ss.; J. BARRAT, *op. cit.*, 71.

<sup>32</sup> J. BARRAT, *op. cit.*, 261.

<sup>33</sup> P. IACCI, *Sotto il segno dell'ignoranza*, Egea, 2021; R. BURIONI, *La congiura dei somari. Perché la scienza non può essere democratica*, Rizzoli, 2017.

strare le differenze e le sfumature tra concetti e azioni»<sup>34</sup>. Diversamente, disattendere il «programmatico aggiornamento critico della cultura degli individui e delle collettività» significherebbe il generarsi di una nuova ignoranza di massa che contribuirebbe al moltiplicarsi in misura esponenziale del già elevato numero degli idioti, «nel senso greco del termine, ossia di persone incapaci di partecipare con una sufficiente consapevolezza alla vita politica e culturale, perché chiuse nella particolarità della loro professione e nei limiti della loro esistenza e dei loro immediati interessi»<sup>35</sup>. Il rischio connesso è quello della polarizzazione e concentrazione del sapere e del potere al vertice della gerarchia sociale, riservandosi l'accesso e la gestione degli algoritmi e delle banche dati ad una oligarchica cerchia di privilegiati, e il confinamento del resto dei consociati in condizioni di ignoranza e povertà intellettuale<sup>36</sup>. La diffusione capillare delle conoscenze scientifiche si trova, in ultima istanza, in un rapporto di diretta ed immediata funzionalizzazione rispetto all'esercizio della sovranità mediante istituti di democrazia diretta ed indiretta: la partecipazione informata e «il buon funzionamento di una società democratica pervasa dall'iperspecializzazione, regolata dal principio di maggioranza e al tempo stesso rispettosa del pluralismo dei valori, è legato indissolubilmente alla probabilità che le nostre opinioni siano corrette o vere»<sup>37</sup>. *Le XXVI Lezioni su Intelligenza Artificiale e diritto* che seguono apportano un contributo essenziale ai fini descritti.

4. – Quali dunque le ragioni giustificative della fiducia nei confronti della scienza? Tra le tante, alcune sembrano meritevoli di apprezzamento particolare.

La prima ragione giustificativa assume una connotazione puramente applicativa e deriva dalla percezione degli aspetti funzionali della scienza, indipendentemente dall'orientamento politico di fondo. Di tali aspetti ci fidiamo «perché li sperimentiamo direttamente ogni giorno: indirettamente, e in modo irriflesso, ci fidiamo anche degli scienziati che con le loro teorie li hanno resi possibili»<sup>38</sup>.

La seconda ragione giustificativa deriva dalla percezione dell'unanimità del consenso che regna in seno alla comunità degli scienziati e della intersoggettività delle conoscenze scientifiche, che inevitabilmente rafforza la persuasione della correttezza delle teorie scientifiche condivise dalla comunità degli esperti e della buona fede che le caratterizza<sup>39</sup>.

La terza ragione è di ordine oggettivo e deriva dal realismo scientifico. La capacità della scienza di descrivere un mondo indipendente dai valori, dagli interessi, dalle esigenze umane, in modo indipendente dalla mente, cogliendo aspetti che prescindono dalla presenza dell'uomo sulla Terra, prova che «le teorie scientifiche sul mondo naturale corrispondono, almeno approssimativamente, a 'come stanno veramente le cose' indipen-

<sup>34</sup> R. BODEI, *Dominio e sottomissione. Schiavi, animali, macchine, Intelligenza Artificiale*, cit., 328.

<sup>35</sup> *Ibidem*, 329-330.

<sup>36</sup> *Ibidem*, 330.

<sup>37</sup> M. DORATO, *Disinformazione scientifica e democrazia. La competenza dell'esperto e l'autonomia del cittadino*, Raffaello Cortina, 2019, 106-107.

<sup>38</sup> *Ibidem*, 81.

<sup>39</sup> *Ibidem*, 81.

dentemente da come vorremmo che stessero e quindi dagli arbitri e dagli interessi soggettivi dei singoli scienziati. Questa forma di oggettività forte sarebbe quindi capace di spiegare perché esiste consenso tra gli scienziati (l'oggettività più debole o la validità intersoggettiva) senza chiamare in causa fattori esclusivamente sociologici»<sup>40</sup>. La scienza non può essere identificata con uno strumento di oppressione patriarcale, né con un costruito sociale: «*it's simply true. Or at least truth is real and science is the best way we have of finding it*»<sup>41</sup>.

La quarta ragione ha una radice procedurale. Se, ripercorrendo la storia e la filosofia della scienza degli ultimi due secoli, può mettersi in dubbio l'esistenza di un unico, aureo metodo scientifico, ciò non può significare simultaneamente rinunciare ad una difesa della scienza dai suoi detrattori. La superiore affidabilità delle tesi scientifiche deriva dal processo sociale che le produce. Tale processo non è perfetto, ma la nostra fiducia non deve andare agli scienziati – per quanto saggi o autorevoli possano essere – ma alla *scienza in quanto processo sociale*, in quanto raggiunge e garantisce il suo consenso solo dopo avere sottoposto le proprie tesi a uno scrutinio rigoroso e plurale, in seguito ad un dibattito che trova la propria naturale sede di svolgimento in seno alla comunità scientifica, formata da esperti e specialisti dei diversi settori<sup>42</sup>. All'apparente connotazione materialistica e meccanicistica che sembra avvolgere il determinismo scientifico è in realtà sotteso un processo dialettico partecipato e condiviso, preordinato alla individuazione di risultati e regole condivise, cui contribuiscono esperti che sottopongono le proprie posizioni all'approvazione o critica dei membri della comunità cui appartengono. Ciò comporta una rilevantissima responsabilizzazione degli scienziati nei confronti della società contemporanea, cui deve accompagnarsi la preservazione, nella massima misura possibile, della loro autonomia di fronte alle pressanti esigenze sociali, pressioni politiche, spinte economiche<sup>43</sup>.

La salvaguardia di tale autonomia, d'altro canto, non può ovviamente tradursi in un altezzoso e superbo isolamento della scienza, i cui apporti – ad esempio l'intelligenza artificiale – non possono non essere compatibilizzati con le peculiari pretese dei diversi ambiti della società, *in primis* quello del diritto e dell'etica giuridica, che comprensibilmente reclamano l'arricchimento del *software* delle macchine, in via complementare ed integrativa, mediante principi condivisi e diffusi in seno alla società medesima. Non esiste, in realtà, un'alternativa quale soggetto della storia tra uomo e tecnica, ma si impone una convivenza tra uomo e tecnica, oltre che con ulteriori fattori (ambiente, clima, flora, fauna), nel contesto di un *koinocene* o *novacene* policentrico che, rifuggendo da un ormai obsoleto antropocentrismo, veda l'uomo interagire in un rapporto di equilibrio con i fattori medesimi. Il dialogo tra scienza e società implica necessariamente, d'altra parte, l'affidabilità degli interlocutori che, se sul versante della scienza è garantita dal processo dialettico in precedenza segnalato, presuppone su altro versante un lungo e capillare processo di al-

<sup>40</sup> *Ibidem*, 84.

<sup>41</sup> R. DAWKINS, *Science is not an instrument of patriarchal oppression*, in *The Spectator*, 13.3.2021.

<sup>42</sup> N. OREKSES, *Perché fidarsi della scienza?*, Bollati Boringhieri, 2021.

<sup>43</sup> G. GOBO, V. MARCHESELLI, *Sociologia della scienza e della tecnologia*, Carocci, 2021; M. BUCCHI, *Scienza e società. Introduzione alla sociologia della scienza*, Raffaello Cortina, 2020; A. CERRONI, Z. SIMONELLA, *Sociologia della scienza*, Carocci, 2014.

fabetizzazione della società quanto ai profili essenziali che connotano, nel caso di specie che ci riguarda, l'intelligenza artificiale e le sue attuali e potenziali applicazioni, integrando la formazione scientifica con quella umanistica<sup>44</sup>. Su un primo versante, l'esigenza di "caricare" con sufficiente efficacia e sicurezza i valori e principi etici, politici, giuridici nelle macchine intelligenti risulta soddisfatta mediante la predisposizione di codici etici e di database per i dispositivi intelligenti che somministrano loro milioni di situazioni, immagini, simulazioni, casi<sup>45</sup> e consentano da parte dei dispositivi medesimi l'adozione di decisioni massimamente conformi ai valori e principi stessi<sup>46</sup>. Su un secondo versante – relativo alla conoscenza dei dispositivi intelligenti e delle concrete ovvero potenziali applicazioni dei medesimi – è invece imprescindibile una fondamentale opera di drenaggio culturale che in modo capillare diffonda in seno alla società e a livello di tutte le sue articolazioni la consapevolezza dei consociati circa la realtà multiforme e poliedrica dell'Intelligenza Artificiale e la sua funzionalizzazione al miglioramento della condizione umana: «oltre al diretto intervento sui processi in corso da parte delle istituzioni o alla responsabilità dei tecnici, degli scienziati, degli industriali, degli uomini di cultura e dei cittadini, l'educazione è lo strumento migliore per abbreviare queste nuove "doglie del parto"»<sup>47</sup>. Oltre che capillare, tale opera di diffusione delle conoscenze deve essere costante e ciclica, reiterata nel tempo: «data la veloce obsolescenza delle nostre informazioni e delle stesse macchine, occorre introdurre urgentemente il sistema della *continuing education* o del *long life learning*, inventando dei sistemi educativi che prevedano (...) il periodico richiamo dei cittadini all'aggiornamento delle loro conoscenze e della cultura generale, e un addestramento, sul piano lavorativo, ai *Cyber Physical Systems*, e quindi alla capacità di operare in processi che connettono il lavoro umano alle nuove tecnologie, in modo tale che il primo stia alla guida, e non al traino delle seconde, così che gli uomini non diventino protesi stupide di macchine intelligenti»<sup>48</sup>. Si tratta del graduale processo formativo di una *Bildung*, nel senso hegeliano del termine, che deve investire non solo i rappresentati, ma anche i rappresentanti in seno alle moderne democrazie: «la democrazia basata sul principio di maggioranza, che esprime la volontà del popolo, funziona solo a condizione che valga il principio di competenza»<sup>49</sup>. Si tratta di leggere il nostro tempo come età della conoscenza e di indirizzare la cultura democratica alla pratica diffusa e organizzata della comunicazione della scienza, intesa come orientamento principe per sviluppare gli aspetti positivi della nostra epoca con la trasformazione dell'informazione in conoscenza, per costituire una cittadinanza scientifica<sup>50</sup>. L'identificazione del gruppo o rappresentante cui delegare il compito di fornire risposte tecniche corrette implica «la necessità di un'alphabetizzazione scientifica e di un più elevato livello culturale, in modo che le probabilità che i cittadini

---

<sup>44</sup> M. DORATO, *Disinformazione scientifica e democrazia*, cit., 105 ss.

<sup>45</sup> M.C. CARROZZA, *I robot e noi*, il Mulino, 2017, 24 ss.

<sup>46</sup> R. BODEI, *Dominio e sottomissione. Schiavi, animali, macchine, Intelligenza Artificiale*, cit., 322-323.

<sup>47</sup> *Ibidem*, 326.

<sup>48</sup> *Ibidem*, 327.

<sup>49</sup> M. DORATO, *Disinformazione scientifica e democrazia*, cit., 110.

<sup>50</sup> P. GRECO, *Homo. Arte e scienza*, Di Renzo, 2021 e già N. WIENER, *Introduzione alla cibernetica. L'uso umano degli esseri umani*, Bollati Boringhieri, 2012.

assegnano alle varie ipotesi siano più vicine al vero»<sup>51</sup>. Diversamente si produce una verticalizzazione del processo decisionale a favore di un'oligarchia inconsapevole e ottusa (rappresentanti disinformati) e incompetente (scienziati venditori di fumo e ciarlatani) ovvero la resa e la capitolazione a favore di un tecnoliberismo viepiù impetuoso<sup>52</sup>.

Non possiamo astenerci dal rilevare come la società contemporanea sia frequentemente percorsa dalla patologica tendenza a discutere su determinati termini senza avere chiaro quale sia il loro vero significato. Lo psittacismo, correlato da Leibniz ai cosiddetti pensieri ciechi o sordi, porta spesso l'uomo a pensare e ragionare utilizzando termini e pensieri formulati da altri, o i suoi stessi ricordi, ripetendoli senza esserne veramente consapevole e senza averli debitamente analizzati<sup>53</sup>. La degenerazione del fenomeno è evidente ove tale conformazione avvenga sulla base di parametri offerti da *influencer*, *youtuber*, da certa stampa o da certa politica: «la rete e i social network, dando sfogo al narcisismo che porta a pensare che tutti si possano esprimere su tutto, non infrequentemente propagano come un potente virus proprio la teoria sbagliata, diffondendo rapidamente credenze del tutto erronee. La disintermediazione operata da Internet favorisce l'eliminazione parziale o totale di qualunque filtro tra esperti/scienziati e cittadini/ineperti, cosicché la voce dei primi perde qualsiasi autorità»<sup>54</sup>.

Infine, ragione che induce alla fiducia nei confronti della scienza consiste nell'attitudine libertaria della tecnica e nello sprigionarsi della sua potenza a fini di affrancamento, emancipazione, riscatto: «la potenza della tecnica libera» e «crea le condizioni perché il pensiero possa liberarsi, e concepire l'umano nella sua integrità e nell'infinita potenzialità racchiusa nella sua finitezza»<sup>55</sup>. Se gli spazi dischiusi da tecnica e potenza possono trovare concretizzazioni e realizzazioni diverse, finanche estreme, declinazioni capaci di tradursi nell'annientamento dell'umano ovvero nella sua piena realizzazione, la scelta fra le diverse vie «dipende dal rapporto che si instaura fra il livello di progresso tecnico raggiunto in una determinata epoca, e la capacità diffusa – sociale, politica, istituzionale, morale – di gestire la potenza conseguita orientandola in modo da avvicinarsi il più possibile a una condizione ideale dove essa possa venire impiegata per proteggere nel suo insieme l'integrità dell'umano – di tutto l'umano nella sua granulare molecolarità – e accrescere in modo omogeneo i suoi livelli di auto-riconoscimento e di autonomia»<sup>56</sup>. La scienza e la tecnica sono veicolo di libertà ed eguaglianza. Il verso del cammino della storia è rivolto, hegelianamente, nella direzione della libertà, «indicata dalla tendenza a raggiungere, da parte di fasce sempre più ampie rispetto alla totalità dell'umano, il rapporto più alto possibile fra la potenza tecnica disponibile – e dunque il controllo sulle proprie condizioni di esistenza

---

<sup>51</sup> *Ibidem*, 111.

<sup>52</sup> S. VACCARO, *Gli algoritmi della politica*, Elèuthera, 2021; E. SADIN, *La silicolonizzazione del mondo. L'irresistibile espansione del liberismo digitale*, Einaudi, 2018; E. SEVERINO, *Il tramonto della politica*, Rizzoli, 2017.

<sup>53</sup> G.W. LEIBNIZ, *De arte combinatoria*, 1666.

<sup>54</sup> M. DORATO, *Disinformazione scientifica e democrazia*, cit., 101.

<sup>55</sup> A. SCHIAVONE, *Progresso*, cit., 108.

<sup>56</sup> A. SCHIAVONE, *Progresso*, cit., 114.

– e il riconoscimento e la valorizzazione della propria esistenza»<sup>57</sup>. In tal senso, «il progresso tecnico definisce la forma generale di *tutta* la storia dell'umano, nell'infinita varietà dei suoi aspetti particolari. Ciò non vuol dire che la spinta in avanti assicurata da questo avanzamento si trasferisca meccanicamente da un piano all'altro, e che tutto proceda insieme allo stesso tempo. Ma significa che l'intera storia riflette e rielabora in ogni sua parte il progresso della tecnica, e ne costituisce per così dire un'interpretazione, che può condurre in direzioni differenti: verso l'abisso o verso una realizzazione sempre più compiuta dell'umano. La scelta dipende dal rapporto, che ogni volta si realizza, fra controllo e potenza, fra ragione e dominio. E si tratta di una relazione che, sui tempi lunghi, ha teso finora sempre più a riequilibrarsi dalla parte della ragione»<sup>58</sup>.

5. – L'uomo ha sperimentato il passaggio dall'*homo faber* classico (capace di dominare la natura attraverso la propria intelligenza) all'*homo creator*, ha concepito strumenti in grado di distruggere il proprio mondo che non è in grado di controllare<sup>59</sup>, prova la “vergogna prometeica” percependo la propria subalternità, in quanto novello Prometeo, al mondo delle macchine da lui stesso create, avverte un senso di “dislivello”, di non sincronicità, tra sé e i propri prodotti meccanici che, sempre più nuovi ed efficienti, lo oltrepassano, facendo sì che si senta “antiquato”<sup>60</sup>. Oltre che perfetta la macchina è instancabile, ripetibile, standardizzata, riproducibile, e possiede una specie di eternità che all'individuo umano è negata: ne deriva una rivalità, una impari gara dell'uomo, una inversione dei mezzi con i fini, che rende evidente come con l'avanzamento della tecnica l'uomo stia mettendo in pericolo la sua esistenza<sup>61</sup>. Se l'uomo, peccando di un ingenuo antropocentrismo, credeva di poter dominare la natura attraverso la tecnica, ora la situazione è rovesciata. Non è più l'uomo il soggetto della storia, bensì la tecnica. Si è resa definitivamente evidente la dirompente dissociazione «fra la velocità dell'avanzamento della tecnica da un lato, e la fatica o l'incapacità del resto della storia umana nell'adeguarsi in modo positivo, dall'altro»<sup>62</sup>. Una sconnessione, in altri termini, «che ha visto da una parte l'accumularsi degli sviluppi impetuosi dell'intelligenza tecnologica e scientifica, e della sua proiezione trasformatrice sulla realtà di ogni singola vita; e, dall'altra, ha assistito al prodursi corrispondente di una sempre maggiore difficoltà nel metter in campo una progettualità culturale e sociale, e una razionalità politica e di governo – sia nell'ordine geopolitico, sia all'interno dei singoli Stati – appena proporzionata agli scenari che si andavano determinando»<sup>63</sup>. Una dissociazione ed uno squilibrio strutturali e sistemici tra potenza e ragione. L'af-

---

<sup>57</sup> *Ibidem*, 117.

<sup>58</sup> *Ibidem*, 118-119.

<sup>59</sup> E. PULCINI, *Dall'homo faber all'homo creator. Scenari del post-umano*, in I. SANNA (ed.), *La sfida del post-umano. Verso nuovi modelli di esistenza?*, Studium, 2005, 13 ss.

<sup>60</sup> G. ANDERS, *L'uomo è antiquato*, Bollati Boringhieri, 2007.

<sup>61</sup> G. ANDERS, *Il mondo dopo l'uomo*, Bollati Boringhieri, 2008.

<sup>62</sup> A. SCHIAVONE, *Progresso*, cit., 32.

<sup>63</sup> A. SCHIAVONE, *Progresso*, cit., 33.

fanno della storia nel rincorrere in modo razionale la tecnica decollata con esponenziale ed inarrestabile velocità.

Ma non si tratta solo della divaricazione tra capacità di produrre e previdenza degli effetti futuri: come osserva L. Grion, «si sta ampliando la scissione tra la nostra capacità di pensare razionalmente i rischi possibili e la possibilità di sentirli realmente tali a livello emozionale. Ciò comporta che, pur *conoscendo* i problemi, non riusciamo a *sentirli* tali; e dunque non ci attiviamo come dovremmo per controllare gli effetti prevedibili delle nostre azioni. È come se non fossimo adeguatamente attrezzati alle sfide che lo sviluppo tecnologico ci ha poso di fronte»<sup>64</sup>. È l'uomo in cui H. Jonas vorrebbe vedere instillato il timore nei confronti di un potere che potrebbe sfuggirgli di mano, disposto ad un bagno di umiltà che gli permetta di identificare i pericoli incombenti e le soluzioni finalizzate ad arginarli, un'euristica della paura quale antidoto contro la *hybris* del nuovo millennio<sup>65</sup>.

Non si tratta tuttavia unicamente, a nostro avviso, di un deficit di umiltà, pregio che raramente ricorre nella natura umana: l'uomo appare piuttosto, nella fase attuale di evoluzione, disorientato, smarrito, sperduto, spaesato di fronte alla finitudine che di tutte le cose la scienza ha svelato: dell'Universo, della Terra, delle specie, di ognuno di noi<sup>66</sup>. È una "sindrome da futuro" generata da un deficit di cultura e di governo cui non sembra in grado di porre rimedio la politica, «ridotta anch'essa, quasi dovunque, a una sfida di corto respiro, impoverita nel linguaggio e nelle idee, a sua volta schiacciata perennemente sul quotidiano; una competizione fra contendenti che si rassomigliano in una sempre maggiore e reciproca identificazione al ribasso, senza progetto, senza talenti, senza visioni del mondo, senza prospettive diverse che non siano la conquista o la conservazione pura e semplice del potere, qui e ora – e che il resto aspetti»<sup>67</sup>. Feste, farina e forza di memoria borbonica: l'assistenzialismo demagogico e deresponsabilizzante accompagnato dal giustizialismo e dalla evaporazione dello Stato di diritto, quanto meno nell'attuale esperienza italiana.

L'uomo ricerca un senso dell'esistenza accettando la propria finitezza e affidandosi alla scienza, consegnandole la propria perpetuazione. L'umano crea il post-umano consapevole della propria parzialità e imperfezione: l'uomo che ha visto dissolversi l'ologramma delle ideologie utopistiche, che avevano promesso il riscatto escatologico dei suoi destini ovvero che hanno disatteso le aspettative di benessere diffuso e di maggiore eguaglianza<sup>68</sup>; l'uomo che ha visto evaporare l'ologramma della religione, che aveva costruito nella speranza, parimenti andata delusa e inevasa, dell'affrancamento, dell'emancipazione, liberazione e redenzione dalla propria condizione di precarietà e transitorietà<sup>69</sup>; l'uomo contemporaneo costruisce il nuovo ologramma dell'Intelligenza Artificiale quale strumento di rigenerazione e che, in una prospettiva non troppo lontana, sia capace di renderlo eterno e

<sup>64</sup> L. GRION, *Chi ha paura del post-umano?*, Mimesis, 2021, 29.

<sup>65</sup> H. JONAS, *Tecnica, medicina ed etica. Un'etica per la civiltà tecnologica*, Einaudi, 1977.

<sup>66</sup> T. PIEVANI, *Imperfezione*, Raffaello Cortina, 2019; ID., *Finitudine*, Raffaello Cortina, 2020.

<sup>67</sup> A. SCHIAVONE, *Progresso*, cit., 45.

<sup>68</sup> M. CACCIARI, P. PRODI, *Occidente senza utopie*, il Mulino, 2016.

<sup>69</sup> *Dio è morto* secondo F. NIETZSCHE, in *Così parlò Zarathustra*, Adelphi, 1986 e *La gaia scienza*, Adelphi, 1977.

immortale. Una sfida a Dio?<sup>70</sup> Una forma di tecno-paganesimo?<sup>71</sup> Una nuova “religione” quale sistema capace di dare senso all’esperienza di un mondo smaterializzato in cui l’invisibile e l’intangibile divengono decisivi?<sup>72</sup> Un paradigma di orientamento in un mondo in vertiginosa accelerazione, percepito tuttavia come capace di distruggere tutto ciò che amiamo, ove l’uomo prova un misto di entusiasmo e orrore, esaltazione e angoscia, accettazione e ripulsa dinanzi al procedere della storia?<sup>73</sup> Il tentativo di estirpare il germe innato della morte avvertito da Hegel?<sup>74</sup> Si tratta comunque di un approccio che denota il superamento da parte dell’uomo della percezione della immodificabilità dei propri limiti e della propria fragilità a favore di una nuova e diversa concezione che li considera come meri confini ora valicabili. La fuga dal presente è affidata alla rivoluzione digitale, a quella neuro-scientifica, a quella genetica e a quella nanotecnologica: «l’immaginario dell’uomo nuovo tecnologico ripropone, in ultima analisi, antiche dinamiche antropologiche di cui proprio la cifra idolatrica sembra coglierne con sorprendente efficacia la logica interna»<sup>75</sup>. L’esito positivo della lotta che l’uomo ingaggia con i propri limiti si traduce – nella narrazione post-umanista – nella conquista e nel «saccheggio del paradiso»<sup>76</sup>, descrivendosi tale successo quale “rapina del cielo” in quanto condizione celestiale non ottenuta per grazia divina, ma raggiunta per opera dell’uomo. Il riemergere della cultura gnostica nel cuore della modernità<sup>77</sup> declina «un profondo senso di insoddisfazione per la situazione presente a cui si accompagna, secondo elemento, la convinzione che le difficoltà dell’oggi possano essere attribuite al fatto che il mondo ha una struttura intrinsecamente deficitaria»<sup>78</sup>. Svuotati di significato e valore umano, «il limite, la fragilità, la vulnerabilità non appaiono che prodotti imperfetti che chiedono d’essere redenti e trasfigurati attraverso un’opera di progressivo perfezionamento»<sup>79</sup>, l’opera di trascendimento continuo del limite apparendo quale unico fattore significativo. La prospettiva dell’oltre-uomo appare in sintesi caratterizzata da quattro profili patologici, che sono stati efficacemente compendiate da chi ha inteso individuarvi «un rapporto malato col tempo e con lo spazio: uno stare al mondo che si rivela incapace di uno sguardo pacificato col proprio presente, guardato solo nei suoi aspetti negativi; (...) un rapporto deformato con la conoscenza tecno-scientifica, considerata non come strumento al servizio della fioritura dell’umano, bensì eletta a fine ultimo in ordine al quale tutto ciò che è possibile, in quanto conoscibile, deve essere altresì fattibile;

---

<sup>70</sup> R. DAWKINS, *Diventare più grandi di Dio*, Mondadori, 2019.

<sup>71</sup> E. DAVIS, *Technopagans. May the Astral Plane Be Reborn in Cyberspace?*, in *Wired*, 1995, 128; ID., *Technognosis: Myth, Magic, and Mysticism in the Age of Technology*, Three Rivers Press, 1999.

<sup>72</sup> M. VENTURA, *Nelle mani di Dio*, il Mulino, 2021; P. SLOTERDIJK, *Dopo Dio*, Raffaello Cortina, 2018.

<sup>73</sup> M. BERMAN, *Tutto ciò che è solido svanisce nell’aria*, il Mulino, 2012.

<sup>74</sup> G.W.F. HEGEL, *Enciclopedia delle scienze filosofiche in compendio* (1817), Bompiani, 2000.

<sup>75</sup> L. GRION, *Chi ha paura del post-umano?*, cit., 210.

<sup>76</sup> A. VACCARO, *L’ultimo esorcismo. Filosofie dell’immortalità terrena*, EDB, 2009, 93 ss.

<sup>77</sup> E. VOEGELIN, *Il mito del mondo nuovo. Saggi sui movimenti rivoluzionari del nostro tempo*, Rusconi, 1976, 19 ss.; H. JONAS, *Il principio gnostico*, Morcelliana, 2011.

<sup>78</sup> L. GRION, *Chi ha paura del post-umano?*, cit., 218.

<sup>79</sup> L. GRION, *Chi ha paura del post-umano?*, cit., 219.



(...) un rapporto deformato con il corpo, rifiutato nei suoi aspetti di fragilità e di corrotibilità e oggetto di potenziamento indefinito così da adeguare la propria attrezzatura fisica alla perfezione a cui l'uomo-nuovo si sente vocato; (...) un rapporto deformato con l'autonomia, assolutizzata e posta a baluardo della libertà umana»<sup>80</sup>.

La categoria del post-umano si fonda «sulla constatazione che l'Uomo – l'*Homo sapiens sapiens* – non possiede più l'appannaggio esclusivo di quella facoltà che ha costituito, *ex origine*, il connotato della sua specificità, e cioè il pensiero»<sup>81</sup>. La conoscenza e l'intelletto, che costituivano il fattore di differenziazione dell'uomo da tutto il resto del cosmo, non rappresentano più una sua esclusiva facoltà. La condivisione con la macchina delle capacità cognitive comporta l'estinzione del rapporto di strumentalità tra la macchina stessa e l'uomo e il dissolversi della supremazia del secondo sulla prima, emergendo un diverso tipo di relazione, improntata a pari-ordinazione se non a autonomia e sovranità della macchina rispetto all'uomo<sup>82</sup>. I *cyborg*, gli organismi cibernetici, «danno l'idea, nella loro articolata casistica, di quanto avanzato sia l'utilizzo di tecnologie infiltrative/ibridative e diffusa la produzione sempre più massificata di corpi “meticci”, assemblaggio di parti eterogenee, metà organismo e metà meccanismo, che cancellano l'identità originaria del corpo umano (la perfetta autosufficienza dell'uomo vitruviano come misura del mondo) per inaugurare la stagione di un corpo “aperto” e “mutante”, espressione di una “tecnologia incarnata” (un *soma* ormai inseparabile e indistinguibile dalla *tekne*)»<sup>83</sup>. L'irradiarsi planetario della scienza e della tecnica implica la revisione dei concetti di individuo, identità, libertà, salvezza, verità, senso, scopo, nonché di quelli di natura, etica, politica, religione, storia, di cui si nutriva l'età umanistica e che ora devono essere riconsiderati, dismessi o rifondati alle radici<sup>84</sup>.

L'antropocentrismo di matrice tolemaica che ha innervato l'umanesimo classico e moderno porta il fardello di gravissime responsabilità, tra l'altro, sul versante della devastazione dell'ambiente nelle sue diverse componenti (fauna, flora, acqua, aria, terra), del clima, delle culture minoritarie. Si impone in tal senso una ricollocazione dell'uomo in un mondo plurale, in seno ad un creato policentrico, multipolare, ri-assegnandogli una posizione paritaria rispetto ad altre creature e ad altri elementi vitali con cui è chiamato a convivere e coabitare e che deve cessare di dominare e possedere: «un'*Humanitas* capace di prendersi cura non solo dell'uomo nel senso limitato della sua individualità di persona o di specie, ma dell'intera catena dell'essere nel mondo: le altre specie viventi, animali e vegetali, le generazioni future, l'habitat, l'oggettività che ci circonda ma che per questo nell'interazione si soggettivizza, i luoghi e le cose (...). Un'*Humanitas* ibrida e ibridante, connettiva e ricombinante, aperta e plurale come ibrida, ricombinante e plurale è la condizione post-umana in cui viviamo le nostre vite»<sup>85</sup>.

<sup>80</sup> L. GRION, *Chi ha paura del post-umano?*, cit., 221-222.

<sup>81</sup> M. REVELLI, *Umano. Inumano. Postumano. Le sfide del presente*, cit., 102.

<sup>82</sup> U. GALIMBERTI, *Heidegger e il nuovo inizio. Il pensiero al tramonto dell'Occidente*, Feltrinelli, 2020.

<sup>83</sup> M. REVELLI, *Umano. Inumano. Postumano. Le sfide del presente*, cit., 114.

<sup>84</sup> U. GALIMBERTI, *Psiche e technè. L'uomo nell'età della tecnica*, Feltrinelli, 2016.

<sup>85</sup> M. REVELLI, *Umano. Inumano. Postumano. Le sfide del presente*, cit., 128.

Se l'umanesimo e l'illuminismo – pur identificabili quali stagioni di forte espansione dell'autonomia e della libertà – hanno espresso l'imporsi di un universale astratto sul particolare reale, la dittatura dell'identico sulla varietà vitale dei molteplici<sup>86</sup>, il *post-umanesimo* implica il congedo dall'eredità dell'uomo cartesiano, soggetto unitario, razionale, titolare di diritto e rapace proprietario di un mondo in cui agisce da signore incontrastato<sup>87</sup>, nonché la definizione di una nuova antropologia fondata sul connubio uomo-macchina<sup>88</sup>: si tratta di concetto che abbraccia le diverse posizioni che condividono il comune denominatore della critica all'umanità tradizionale e dell'affermazione della necessità del superamento dell'idea abituale di uomo, in quanto rigetto di ogni forma di antropocentrismo (*postumanismo*) ovvero in quanto auto-trascendimento umano concepito al centro della storia e convalida del protagonismo del soggetto moderno (*transumanesimo*)<sup>89</sup>.

Cosa attende l'uomo quando raggiungerà la condizione post-umana avendo completato l'attraversamento dei tre ponti immaginati da Kurzweil e Grossman e consistenti nella conservazione di una buona salute fisica e mentale, nella rivoluzione biotecnologica e nel pieno sviluppo delle nanotecnologie e dell'intelligenza artificiale?<sup>90</sup>

Quale il "posto" ed il rilievo dell'uomo al tempo del *Koinocene*, neologismo che tende ad evocare l'epoca delle connessioni o partecipazioni, che impone il riconoscimento in via definitiva dei profondi legami, delle connessioni, delle partecipazioni tra gli esseri viventi e non viventi che coabitano nel pianeta, tra i collettivi umani e quelli non umani?<sup>91</sup> Quale il rapporto dell'uomo con le macchine intelligenti e quale il destino della Terra nella fase di superamento dell'*Antropocene*, era di profonde ferite inferte dall'uomo alla natura, al clima, all'ambiente — l'era geologica in cui la nostra specie si

<sup>86</sup> L. GRION, *Chi ha paura del post-umano?*, cit., 34.

<sup>87</sup> R. BRAIDOTTI, *Il postumano. La vita oltre l'individuo, oltre la specie, oltre la morte*, DeriveApprodi, 2014.

<sup>88</sup> D.J. HARAWAY, *Manifesto cyborg. Donne, tecnologie e biopolitiche del corpo*, Feltrinelli, 2018.

<sup>89</sup> Tali versioni applicative del post-umanesimo sono state compendiate da A. ALLEGRA, *Visioni transumane. Tecnica, salvezza, ideologia*, Orthotes, 2017, 9 ss.; L. GRION, in *Chi ha paura del post-umano?*, cit., 48, propone un sincretismo tra *post-umanesimo* e *post-umanismo* utilizzando i due termini come sinonimi «per indicare il possibile (e auspicato) esito di un processo di progressiva presa di congedo dai limiti dell'attuale condizione umana»; il post-umanesimo, in tal senso, viene concepito «come esito di un cammino che condurrà l'uomo a vincere la sfida ingaggiata coi propri limiti costitutivi», nella persuasione «che la tecnica abbia portato l'uomo alle soglie di un passaggio epocale che lo condurrà, nel breve volgere di qualche decade, a veder radicalmente mutate le coordinate tipiche del proprio star al mondo: vivere senza "data di scadenza", eliminando i tratti di fragilità e di vulnerabilità che oggi caratterizzano l'esperienza umana; godere senza sosta emozioni e sensazioni piacevoli, ponendo invece sotto controllo le emozioni negative, i cattivi ricordi, le esperienze sgradite; poter contare su capacità fisiche e cognitive potenziate; incamminarsi senza timori lungo i sentieri dell'intelligenza artificiale (AI) e dell'ibridazione con le macchine».

Le strategie del transumanesimo si trovano originariamente delineate nelle opere di J. HUXLEY, *New Bottles for New Wine*, Chatto & Windus, 1957; P. TEILHARD DE CHARDIN, *L'avvenire dell'uomo*, Il Saggiatore, 1972; F.M. ESFANDIARY, *Are you a transhuman?: monitoring and stimulating your personal rate of growth in a rapidly changing world*, Warner Books, 1989; M. MORE, N. VITA-MORE (eds.), *The Transhumanist Reader*, Wiley-Blackwell, 2013.

<sup>90</sup> R. KURZWEIL, T. GROSSMAN, *Fantastic Voyage: Live Enough to Live Forever*, Rodale Books, 2004.

<sup>91</sup> B. LATOUR, *La scienza in azione. Introduzione alla sociologia della scienza*, Edizioni di Comunità, 1998; ID., *Politiche della natura. Per una democrazia delle scienze*, Raffaello Cortina, 2000.

è dimostrata un fattore critico per l'intero pianeta<sup>92</sup> – verso un *Novacene* caratterizzato dalla interazione e collaborazione tra l'uomo e le macchine? È oggi difficile asseverare l'ipotesi secondo cui in un prossimo futuro nuovi esseri prenderanno forma dall'intelligenza artificiale che noi abbiamo progettato e che penseranno 10 000 volte più velocemente dell'uomo: è tuttavia plausibile che, in quanto esseri iper-intelligenti, sapranno meglio di noi di essere totalmente dipendenti dallo stato di salute del pianeta. Come noi, anche le macchine avranno bisogno del sistema regolatore di Gaia per sopravvivere e dal momento che Gaia dipende dalla vita organica, sarà loro interesse preservarla. Il *Novacene* potrebbe essere l'inizio della conquista dell'intero cosmo da parte di un'intelligenza diffusa, l'alba di un nuovo universo<sup>93</sup>.

Ciò che è sicuro ed incontrovertibile è che la natura e la scienza hanno profondamente relativizzato la visione antropocentrica della realtà, facendo emergere la crisi dell'individuo e il suo posto nel mondo: la crisi attuale costituisce un'occasione imperdibile per ridefinire l'identità collettiva e restituire un senso al mondo<sup>94</sup>.

La sfida epocale che attende l'uomo nei propri rapporti con la scienza e la tecnologia consiste nello sviluppo e nell'accrescimento, accanto alla *competenza* nella costruzione e nel funzionamento di sistemi intelligenti non biologici, della *comprensione* delle utilizzazioni dei sistemi medesimi, in quanto consapevolezza della loro funzionalizzazione ad obiettivi compatibili con i valori, i principi e gli interessi umani<sup>95</sup>. Accantonando isterie e catastrofismi, l'uomo deve concepire la scienza quale strumento necessario per imprimere alle due parallele di libertà ed eguaglianza la torsione che ne permetta la curvatura e la definitiva convergenza verso un punto di sintesi: il sincretismo tra libertà effettiva ed eguaglianza sostanziale.

---

<sup>92</sup> S. PEPPOLONI, G. DI CAPUA, *Geoetica. Manifesto per un'etica della responsabilità verso la Terra*, Donzelli, 2021.

<sup>93</sup> J. LOVELOCK, *Novacene. L'età dell'iperintelligenza*, Bollati Boringhieri, 2019, 79 ss.

<sup>94</sup> P. BLOM, *Il gran teatro del mondo*, Marsilio, 2021.

<sup>95</sup> M. FERRARIS, *Documanità. Filosofia del nuovo mondo*, Laterza, 2021, 200 ss.

CAPITOLO I

**L'INTELLIGENZA ARTIFICIALE: COS'È?**



# I LEZIONE: L'INTELLIGENZA ARTIFICIALE: COS'È?

## DALLA LOGICA AL *DEEP LEARNING*: UNA BREVE RIFLESSIONE SULL'INTELLIGENZA ARTIFICIALE

*Maurizio Gabbrielli\**

SOMMARIO: 1. Introduzione. – 2. Perché adesso? Una breve prospettiva storica e tre risposte. – 3. Cosa è oggi l'intelligenza artificiale. – 4. Alcune implicazioni economiche. – 5. Conclusioni.

**1. Introduzione.** – Assistiamo quotidianamente alla celebrazione delle “magnifiche sorti e progressive” dell'intelligenza artificiale, dalle ultime applicazioni di transfer learning, che ci fanno apparire la cartolina di un canale delle fiandre nello stile del famoso quadro *La Nuit étoilée* del 1889 di Van Gogh<sup>1</sup>, alle applicazioni di manutenzione predittiva, che permettono di prevedere in anticipo i malfunzionamenti di una linea di produzione in un impianto industriale. Certamente l'intelligenza artificiale, da qualsiasi prospettiva la si voglia considerare (e vi sono sicuramente molte prospettive e angolazioni diverse per occuparsene) offre spunti scientifici, tecnologici, epistemologici e filosofici molto importanti, dunque l'interesse diffuso è comprensibile e giustificato.

Tuttavia, nell'entusiasmo generale, iniziano ad emergere anche varie voci di dissenso, e per motivi diametralmente opposti. Da un lato vi sono obiezioni di vecchia data alla possibilità di realizzare effettivamente qualche cosa di artificiale che si possa definire intelligente<sup>2</sup>, obiezioni che hanno alimentato per decenni il dibattito scientifico soprattutto nell'ambito delle scienze cognitive. Di segno opposto invece sono le preoccupazioni di coloro che credono che le macchine, grazie alla continua crescita della loro capacità di accumulare ed elaborare conoscenza, potranno raggiungere un punto di

---

\* Ordinario di Informatica – Coordinatore del corso di Laurea Magistrale in Artificial Intelligence dell'Università di Bologna.

<sup>1</sup> Si veda: GATYS, ECKER, BETHGE, *A neural algorithm of artistic style*, in *CoRR*, 2015, abs/1508.06576.

<sup>2</sup> Si veda ad esempio il famoso esperimento mentale della stanza cinese in SEARLE, *Minds, brains, and programs*, in *Behavioral and Brain Sciences*, 1980, 3. Tale esperimento, secondo l'autore, prova che non è possibile costruire una macchina pensante, ossia nega la possibilità della cosiddetta intelligenza artificiale forte.

“singolarità tecnologica” oltre il quale l’intelligenza umana verrebbe superata da quella artificiale.

Senza la pretesa di voler fornire risposte a questa o altre domande sul futuro dell’intelligenza artificiale, in questa breve nota intendiamo riflettere sullo stato attuale di questa disciplina, su alcune sue implicazioni tecnologiche e scientifiche, ma anche sociali ed economiche, e lo vogliamo fare partendo da una domanda molto semplice: “perché tutto questo accade adesso”? È una domanda la cui risposta non è scontata perché, contrariamente a quello che il profano potrebbe pensare, l’intelligenza artificiale è vecchia tanto quanto l’informatica e, come questa, è più vecchia dei computer, ossia delle macchine che poi ne hanno reso possibile la realizzazione pratica. La maggior parte delle tecniche che oggi usiamo in intelligenza artificiale, reti neurali incluse, nei principi di base hanno un’età di vari decenni. Quindi, perché adesso? Per approfondire la questione e cercare di capire meglio la natura di questa disciplina, seguendo Comte<sup>3</sup>, può essere utile inquadrare la questione in una prospettiva storica.

**2. Perché adesso? Una breve prospettiva storica e tre risposte.** – Secondo una definizione condivisa da molti l’intelligenza artificiale si occupa di realizzare strumenti (software e hardware) che siano capaci di eseguire compiti normalmente associati all’intelligenza naturale. Questa almeno era la visione di coloro che coniarono il termine “artificial intelligence” in occasione del workshop che si tenne a Dartmouth nel 1956 e che viene considerato l’evento “ufficiale” della nascita della disciplina<sup>4</sup>. Tuttavia, il sogno di costruire macchine che emulassero pensieri e comportamenti umani ha origini ben più lontane.

Il sillogismo aristotelico, come noto, fu il primo sistema di argomentazione logica che a partire da delle premesse permettesse di dedurre una conclusione, fornendo così un primo esempio di ragionamento formale (nel senso che la correttezza del ragionamento dipende dalla sua struttura, o forma, e non dal significato dei termini impiegati) e dunque un primo sistema di ragionamento meccanizzabile.

Bisogna però aspettare fino al XIII secolo con Ramon Llull (1232-1316) per avere una prima formulazione esplicita dell’idea di calcolo meccanico. Llull con la sua *Ars combinatoria* (secondo la denominazione di Leibniz) non solo propose un linguaggio logico algebrico per rappresentare la conoscenza e generare nuove verità a partire da premesse note, ma fornì anche una implementazione, come diremmo oggi, del suo modello di calcolo. Llull realizzò infatti una macchina fisica costituita da dischi di carta concentrici che, ruotati opportunamente, permettevano di combinare in vario modo i concetti rappresentati in modo simbolico su di essi mediante lettere. Llull ha avuto una influenza

---

<sup>3</sup> «On ne connaît pas complètement une science tant qu’on n’en sait pas l’histoire», COMTE, *Cours de philosophie positive*.

<sup>4</sup> Nella proposta originaria per il workshop di Dartmouth possiamo leggere: «The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it». MCCARTHY, MINSKY, ROCHESTER, SHANNON, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. August 31, 1955, in *AI Magazine*, 2006, 27, 4.