

XV. L'APOCALISSE DELL'INTELLIGENZA ARTIFICIALE

di GIOVANNI AMENDOLA

1. *L'emergenza (anche) religiosa di un'etica condivisa per l'intelligenza artificiale*

Nell'ultimo decennio, a partire all'incirca dal 2011, grazie alla congiunzione di almeno tre fattori: la velocità di calcolo raggiunta dai computers, la grande quantità di dati accumulata su internet (*Big Data*) e l'avanzamento delle tecniche di apprendimento automatico (*Machine Learning* e *Deep Learning*), si sono rapidamente sviluppati sistemi basati sull'intelligenza artificiale (IA) sempre più sorprendenti nelle capacità di riconoscere immagini, di tradurre il linguaggio parlato in testo scritto, di tradurre un testo da una lingua ad un'altra, di fornire risposte adeguate a domande poste da un utente, di guidare automaticamente un veicolo, di produrre diagnosi mediche, eccetera¹. Tuttavia, fin da subito, è emersa una globale preoccupazione di carattere etico strettamente connessa proprio alle tecniche di IA utilizzate e alla presenza sempre più diffusa di tali sistemi. Tra i principali problemi sollevati, inerenti alle tecniche di apprendimento automatico, quello dovuto all'impossibilità di comprendere le motivazioni esatte che conducono un sistema di IA a produrre un certo risultato, a causa dell'approccio statistico che individua correlazioni profonde e nascoste tra i dati ma non la logica soggiacente, e poi quello del valore dei dati di partenza che, se non adeguatamente vagliati, possono generare soluzioni eticamente discutibili e discriminatorie quando utilizzate per prendere decisioni o fare scelte in ambito sociale².

¹ Cfr. M. MITCHELL, *Artificial Intelligence: A Guide for Thinking Humans*, Macmillan, London 2019.

² Cfr. A. MONTANARI, *Artificial Intelligence and Creativity*, in «Acta Philosophica», 32 (2023), n. 1, pp. 35-54.

Per tali ragioni, a partire dal 2014 una molteplicità di agenzie governative, compagnie private, organizzazioni pubbliche e istituti di ricerca hanno individuato principi basilari e generali per una IA etica. In un articolo apparso sulla rivista *Nature Machine Intelligence* nel 2019, tre ricercatori del Politecnico federale di Zurigo hanno rinvenuto ed analizzato quasi un centinaio di diverse documentazioni a riguardo, identificando i cinque principi maggiormente presenti in tutte le linee guida: trasparenza, giustizia ed imparzialità, non-maleficenza, responsabilità e privacy³.

Accogliendo queste istanze emergenti, anche nella Chiesa Cattolica è cresciuta una particolare sensibilità a tali questioni. Tra gli altri, Paolo Benanti ha sviluppato una riflessione sull'etica dell'IA, parlando della necessità di un'algoetica e di una governance in grado di orientare umanamente queste nuove tecnologie ed evitare i pericoli dell'algocrazia⁴. Così nel febbraio del 2020, attraverso la Pontificia Accademia per la Vita, è stata promossa la *Rome Call for AI Ethics*, che ha prodotto un documento, presumibilmente il primo coinvolgente una istituzione religiosa, in cui vengono sottolineati sei principi per l'algoetica: la trasparenza (*transparency*), secondo cui i sistemi di IA devono essere comprensibili a tutti; l'inclusione (*inclusion*), per cui tali sistemi non devono discriminare nessuno, essendo ogni essere umano di pari dignità; la responsabilità (*responsibility*), in quanto ogni qualvolta un sistema di IA viene rilasciato deve esserci qualcuno che se ne assume la responsabilità; l'imparzialità (*impartiality*), per cui tali sistemi non devono seguire o creare pregiudizi; l'affidabilità (*reliability*), in base alla quale i sistemi di IA devono assicurare un corretto funzionamento; infine, la sicurezza e la privacy (*security and privacy*), affinché tali sistemi siano protetti da attacchi digitali e garantiscano la privacy degli utenti⁵. Dopo circa tre anni, per via delle politiche del periodo pandemico, il 10 gennaio 2023, la *Rome Call* è stata sottoscritta anche da rappresentanti dell'Ebraismo e dell'Islamismo, in modo da tendere verso «un'etica condivisa riguardo alle grandi sfide che si aprono nell'orizzonte dell'intelligenza artificiale»⁶.

³ Cfr. A. JOBIN-M. IENCA-E. VAYENA, *The global landscape of AI ethics guidelines*, in «Nature Machine intelligence», 1 (2019), pp. 389-399.

⁴ Cfr. P. BENANTI, *Oracoli. Tra algoetica e algocrazia*, Luca Sossella, Roma 2018.

⁵ Cfr. V. PAGLIA-R. PEGORARO, *The "Good" Algorithm? Artificial Intelligence Ethics, Law, Health*, Pontifical Academy for Life, Vatican City 2021.

⁶ FRANCESCO, *Discorso ai partecipanti all'incontro "Rome Call" promosso dalla Fondazione RenAIssance*, Vaticano, 10 gennaio 2023.

2. *L'intelligenza artificiale nella crisi antropologica in atto: l'urgenza di una nuova redenzione dell'umano*

La questione etica dell'IA, verso cui convergono le più svariate sensibilità planetarie, presuppone imprescindibilmente un orizzonte ben più vasto, in quanto, come ogni questione etica relativa alla prassi umana, richiede una visione antropologica condivisa ed una direzione verso cui tendere come umanità. Questo, a nostro giudizio, è il vero cuore della problematica, purtroppo ancora scarsamente affrontato a livello globale, mentre le voci culturali più autorevoli, ma ben poco ascoltate, cercano di condurci a riconoscere seriamente il fatto di trovarci in una crisi antropologica senza precedenti⁷, essendo dinanzi e dentro un cambiamento epocale⁸. Papa Francesco, nel discorso in occasione della *Rome Call*, aveva chiaramente affermato che «la “galassia digitale”, e in particolare la cosiddetta “intelligenza artificiale”, si trova proprio al cuore del cambiamento d'epoca che stiamo attraversando»⁹. In questa situazione critica di trasformazione, in cui sembra dissolversi ogni punto di riferimento in grado di offrire una base per orientarsi (il clima relativistico-nichilistico diffuso largamente nelle società occidentali ed occidentalizzate), si finisce inevitabilmente con l'essere dominati dalle logiche planetarie egemoni, ovvero quelle del finanz-capitalismo neoliberista, che sta indebolendo sempre più le democrazie del pianeta¹⁰.

In un tale contesto, l'IA rischia di divenire troppo facilmente un'ulteriore fonte di disuguaglianza tra gli esseri umani, di ingiustizia sociale e di nuove schiavitù. Ci ritroviamo, pertanto, ben distanti dalle posizioni di chi,

⁷ Per un'interpretazione della crisi antropologica in senso evolutivo come passaggio pasquale di liberazione cfr. M. GUZZI, *La svolta. La fine della storia e la via del ritorno*, Jaca Book, Milano 1986 (riedito da San Paolo, Milano 2022).

⁸ Il cambiamento d'epoca di cui parla papa Francesco e a cui ci riferiamo è diverso da quello descritto ad esempio in P. BENANTI, *Digital Age. Teoria del cambio d'epoca. Persona, famiglia e società*, San Paolo, Milano 2020, dove si tratterebbe del passaggio dal paradigma scientifico-causale a quello algoritmico-correlazionale. Per noi il cambiamento d'epoca ha una portata ancora più ampia ed affonda le sue radici nel dinamismo cristologico della storia evolutiva antropo-cosmica.

⁹ FRANCESCO, *Discorso ai partecipanti alla plenaria della Pontificia Accademia per la Vita*, Vaticano, 28 febbraio 2020.

¹⁰ Cfr. FRANCESCO, *Laudato Si'*, 54 e 109; *Fratelli tutti*, 12 e 168; e tra i pensatori contemporanei S. LATOUCHE, *Decolonizzare l'immaginario. Il pensiero creativo contro l'economia dell'assurdo*, EMI, Verona 2004; D. HARVEY, *Breve storia del neoliberismo*, Saggiatore, Milano 2007; L. GALLINO, *Finanzcapitalismo. La civiltà del denaro in crisi*, Einaudi, Torino 2013.

come il professore di informatica del MIT Alex Pentland, teorizza invece un ottimismo esagerato prospettando una città e una società dei dati come semplicemente “più umane”, “più sane”, “più sicure” e “più efficienti”¹¹. Mentre Pentland prospetta tali scenari, da altri versanti vengono infatti posti in luce fenomeni profondamente disumanizzanti connessi all'utilizzo di tali tecnologie. Ci riferiamo nello specifico a tre autori, che, da contesti diversi, giungono a medesime conclusioni, proprio a partire da una riflessione che tiene conto dell'attuale situazione politica e culturale delle società occidentali e non solo.

Il filosofo sudcoreano dell'università di Berlino, Byung-Chul Han, nel saggio *Psicopolitica*, riconosce con chiarezza l'attuale tacito predominio di una fiducia di fondo nella misurabilità e nella quantificabilità della stessa vita umana, connessa all'affidamento, stavolta più esplicito, come in Pentland, proprio a quei sistemi di IA basati e guidati dai big data. La stessa persona umana, in questa visione di fondo, si ritrova ridotta ad un'enorme quantità di dati, a cui aderisce irriflessivamente con una fede cieca. Si tratta di ciò che il filosofo chiama *quantified self*:

«Il corpo viene dotato di sensori che automaticamente registrano dati. Vengono misurati la temperatura corporea, il tasso glicemico, l'assunzione e il consumo di calorie, i tragitti dei singoli o la percentuale di grasso corporeo. Vengono contati i battiti cardiaci durante la meditazione. Persino nei momenti di riposo hanno importanza prestazione ed efficienza. Vengono protocollati anche condizioni mentali, stati d'animo e attività quotidiane: mediante l'automisurazione e l'autocontrollo dovrebbe migliorare la prestazione fisica e spirituale. La mera massa di dati così accumulata non risponde però alla domanda: Chi sono io? Anche il *quantified self* è una tecnica dadaista del Sé, che lo svuota completamente di senso. Il Sé è scomposto in dati fino al vuoto di senso. Il motto del *quantified self* è: *Self knowledge through numbers* (“Conoscenza di sé attraverso i numeri”). Per quanto sterminati possano essere, da dati e numeri non si ricava alcuna conoscenza del sé. I numeri non raccontano nulla del Sé. Contare non è raccontare; il Sé, infatti, deriva da un racconto. Non il contare, ma il raccontare conduce alla scoperta e alla conoscenza di sé»¹².

L'essere umano, consegnato a questa illusoria quantificazione di sé, finisce con l'essere, ancor di più rispetto al passato, una facile preda delle

¹¹ Cfr. A. PENTLAND, *La fisica sociale. Come si propagano le buone idee*, EGEA, Milano 2015.

¹² B.C. HAN, *Psicopolitica. Il neoliberismo e le nuove tecniche del potere*, Nottetempo, Milano 2016, pp. 71-72.

logiche fagocitanti dei mercati globali e delle multinazionali tecnologiche. La sociologa statunitense Shoshana Zuboff, nell'opera *Il capitalismo della sorveglianza*¹³, fa notare come i dati privati incamerati dai grandi colossi informatici come Google, Facebook, Amazon, Microsoft, ecc., vengano venduti ed usati per orientare e dirigere i comportamenti futuri delle persone nel mercato. Questi dati possono mettere in luce dipendenze ed automatismi presenti nella psicologia di ciascuno. Infatti, come messo ben in evidenza dalla psicologia comportamentale, gli esseri umani agiscono quasi sempre in una logica di azione-reazione automatica, mossi da programmi inconsci presenti e radicati nella psiche, come ciascuno può sperimentare attraverso i propri pensieri automatici¹⁴. Pertanto, offrire molti dati su se stessi può permettere ai sistemi di IA basati sui dati di prevedere le possibili risposte umane a determinati stimoli di input, individuando gli output degli algoritmi psichici inconsci, ed, in tal modo, orientare l'essere umano verso i propri obiettivi strategici di incremento del profitto. Lo ha descritto molto bene lo storico israeliano Yuval Noah Harari, che giunge a parlare di un *essere umano hackerabile*:

«Conoscenza biologica moltiplicato per potenza informatica moltiplicato per i dati uguale capacità di hackerare gli esseri umani. Quando le autorità e le aziende avranno sufficienti conoscenze biologiche, sufficiente potenza informatica e sufficienti dati potranno hackerare e manipolare gli esseri umani con modalità senza precedenti. Questo cambierà l'assetto economico, il sistema politico e le nostre vite quotidiane in modi che possiamo appena immaginare, e l'autorità si trasferirà sempre più dagli uomini agli algoritmi»¹⁵.

Gli scenari distopici evidenziati in tanti film di fantascienza, su tutti *Matrix*, sembrano molto prossimi a quelli che abbiamo qui delineato. L'essere umano rischia nuove forme di schiavitù, confinato ad una vita di inconsapevolezza, in una realtà illusoria e virtuale, condizionata da algoritmi che orientano dal subconscio i suoi comportamenti e le sue azioni, riducendo l'essere umano ad un robot o ad un burattino che crede illusoriamente di desiderare liberamente ciò che vuole, mentre è mosso dalle

¹³ S. ZUBOFF, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Luiss University Press, Roma 2019.

¹⁴ Cfr. D. GREENBERGER-C.A. PADESKY, *Penso, dunque mi sento meglio*, Erickson, Trento 2007.

¹⁵ Y.N. HARARI, *21 lezioni per il XXI secolo*, Giunti-Bompiani, Firenze-Milano 2021, p. 423.

potenze egemoniche di questo mondo che ne risucchiano le energie vitali e si consolidano ulteriormente.

Eppure non si tratta di qualcosa di completamente nuovo. Questo condizionamento algoritmico dell'umano trova un legame strettissimo con la descrizione della condizione umana delineata dalle grandi tradizioni culturali e religiose presenti sulla Terra. Innanzitutto, quanto stiamo dicendo, è presente nel cuore stesso della tradizione filosofica occidentale, dove ha trovato espressione, tra le altre, nel mito della caverna di Platone, in cui appunto l'essere umano è concepito in uno stato di prigionia all'interno di una caverna, forzato a guardare sul fondo le ombre proiettate da oggetti che sono alle sue spalle e finendo col ritenere ciò che osserva come la vera realtà. Analogamente, la tradizione ebraico-cristiana ritiene che l'essere umano non vive nella realtà a lui più consona e propria, quella che Dio aveva per lui pensato, ma in un mondo caduto a causa del peccato, ovvero a motivo di una distorsione della libertà che ha prodotto una radicale distorsione del mondo buono creato da Dio stesso. Altrettanto può dirsi delle tradizioni religiose orientali, ad esempio dell'induismo o del buddismo, le quali ritengono che l'essere umano si trovi in una realtà illusoria, ingannato dai sensi e dalla psiche, vivendo in uno stato di ignoranza (*avidyā*), ritenendosi un io separato dal tutto e necessitando di risvegliarsi alla realtà vera.

Sta dunque emergendo nel nostro tempo una nuova modalità di alienazione, quella della riduzione algoritmica dell'umano, che può, al contrario, avviare un'ulteriore fase nella storia di liberazione e di redenzione, attraverso cui giungere ad una comprensione sempre più globale dell'alienazione umana e a delineare ed intraprendere nuovi cammini personali e sociali per sbloccare proprio quei meccanismi psichici automatici che rendono gli esseri umani prigionieri e schiavi del passato e soffocano il potere creativo che agisce in essi¹⁶.

¹⁶ A riguardo ci sembrano molto appropriati i tentativi laboratoriali, avviati da circa venticinque anni, nei gruppi "Darsi pace" ideati dal poeta e filosofo Marco Guzzi, in cui pratiche psicologiche e spirituali di liberazione interiore si congiungono ed integrano con la riflessione culturale e socio-politica (cfr. M. GUZZI, *Darsi pace. Un manuale di liberazione interiore*, San Paolo, Milano 2004; *La nuova umanità. Un progetto politico e spirituale*, San Paolo, Milano 2005).

3. *Entrare nell'intelligenza artificiale per riscoprire l'intelligenza (e la libertà) spirituale*

Riteniamo che, proprio a partire da una conoscenza più adeguata e profonda della stessa intelligenza artificiale, si possa giungere a comprendere che una concezione riduzionistico-algoritmica dell'umano è gravemente deficitaria e rinuncia a dimensioni essenziali dell'essere umano¹⁷.

Una percezione sensibile oltre la sensorialità. Innanzitutto, la sensorialità di un robot (si pensi a telecamere interne o a sensori in grado di recepire suoni o odori) è ben diversa dalla sensorialità umana. Lo descrive acutamente l'inventore del microprocessore, il fisico Federico Faggin, osservando che, dotando un robot di un sensore olfattivo, è possibile fargli riconoscere una rosa attraverso il suo profumo, in quanto le molecole effuse dalla rosa possono essere intercettate da specifici sensori e codificate in sequenze di bit che saranno poi analizzate attraverso particolari reti neurali artificiali, le quali restituiranno come output un'ulteriore sequenza di bit corrispondente al nome "rosa". Il processo è per certi aspetti simile a quanto accade ad un essere umano che, passando vicino ad una rosa, per mezzo dell'epitelio nasale, attraverso le membrane delle cellule olfattive, intercetta le molecole della rosa presenti nell'aria. In tal modo si produce un segnale elettrico che diviene l'input delle reti neurali presenti nella corteccia olfattiva, che produrranno un segnale elettrico a cui sarà associato il nome "rosa". Tuttavia, fa notare Faggin, vi è un discrimine fondamentale nei due processi, in quanto, a differenza del robot, l'essere umano *sente* il profumo della rosa¹⁸.

Una conoscenza oltre la rigorosa descrivibilità. Uno dei padri fondatori dell'intelligenza artificiale, John McCarthy, nel definire tale progetto riteneva che in linea di principio qualsiasi aspetto dell'apprendimento e dell'intelligenza potesse essere descritto in modo talmente preciso e rigoroso da poter costruire una macchina calcolante in grado di simularlo¹⁹. Tuttavia,

¹⁷ Per un'analisi più approfondita cfr. G. AMENDOLA, *Antropo-Logos. La ragione al crocevia di intelligenza artificiale, razionalità scientifica, pensiero filosofico e teologia cristiana*, Studium, Roma 2021, pp. 27-93.

¹⁸ Cfr. F. FAGGIN, *Silicio. Dall'invenzione del microprocessore alla nuova scienza della consapevolezza*, Mondadori, Milano 2019, pp. 182-183.

¹⁹ Cfr. J. MCCARTHY-M.L. MINSKY-N. ROCHESTER-C.E. SHANNON, *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, in «AI Magazine», 27 (2006),

per l'essere umano conoscere o apprendere qualcosa non equivale mai a darne una descrizione minuziosa e precisa. Si pensi ad esempio ad apprendere uno sport o uno strumento musicale oppure a conoscere una nuova pianta o un frutto mai visto prima, come potrebbe essere il mangostano, un frutto diffuso principalmente nel sud-est asiatico. Si potrebbe immaginare che qualcuno, che lo ha assaggiato e lo conosce bene, possa darne una descrizione talmente accurata e precisa da non tralasciare nessun dettaglio circa il suo aspetto o il suo sapore (la descrizione rigorosa richiesta da McCarthy). Ora, se una persona, che non ha mai sentito parlare di quel frutto, leggesse la descrizione fornita, saprebbe indubbiamente molte informazioni a riguardo (ad esempio che ha un sapore per certi aspetti simile ad una pesca e per altri simile ad un litchi), tuttavia quanto appreso dalla descrizione rimarrebbe infinitamente distante rispetto alla conoscenza di colui che lo ha descritto. In altri termini, la conoscenza umana ha sempre a che fare con una esperienzialità indescrivibile sul piano logico-formale, che è invece l'unico piano a cui ha accesso, in ultima analisi, un sistema di intelligenza artificiale²⁰.

Una comprensione oltre la manipolazione sintattica. A partire da novembre 2022 si stanno diffondendo planetariamente sempre più raffinate chatbot basate sull'apprendimento automatico in grado di conversare con un utente umano e di giungere a livelli tali da indurre molti a ritenere di aver superato il noto test di Turing, se non addirittura di essere dinanzi a qualcosa in grado di capire le nostre richieste o di comprendere le nostre emozioni. Eppure ciò che operano questi sistemi di intelligenza artificiale, pensati per l'elaborazione del linguaggio naturale, non è altro che una manipolazione sintattica di simboli senza significato come adeguatamente spiegato dal John Searle attraverso l'esperimento mentale della stanza cinese²¹. Il filosofo statunitense ha immaginato una stanza al cui interno risiede un uomo di lingua inglese che non conosce il cinese, ma che ha a disposizione dei manuali in cui sono presenti delle regole di traduzione che, data una certa sequenza di caratteri cinesi, producono nuove sequenze. Un uomo cinese all'esterno della stanza pone per iscritto delle domande in cinese e le

n. 4, p. 12.

²⁰ Cfr. G. AMENDOLA, *La comprensione dell'"intelligenza" tra intelligenza artificiale, filosofia e teologia*, in «Rivista italiana di filosofia del linguaggio», 17 (2023), n. 1, pp. 163-177.

²¹ Cfr. J. SEARLE, *Minds, Brains and Programs*, in «Behavioral and Brain Sciences», 3 (1980), n. 3, pp. 417-457.

inserisce all'interno. L'uomo inglese utilizza le regole a disposizione per trasformare le sequenze di simboli cinesi in altre sequenze, e così via, finché, alla fine, non ha prodotto un nuovo testo con simboli cinesi, che restituisce all'esterno della stanza. A questo punto, l'uomo cinese, leggendo il testo e comprendendone il significato, ritiene che all'interno della stanza ci sia qualcuno che conosce la lingua cinese. Searle fa notare che, invece, l'uomo inglese ha soltanto applicato un processo meccanico di manipolazione sintattica di simboli senza avere alcuna comprensione del messaggio che ha prodotto. Ciò è quello che accade anche nei moderni sistemi di intelligenza artificiale²².

Una coscienza oltre il cervello. Infine, un ultimo punto su cui vogliamo brevemente soffermarci è il problema filosofico del rapporto mente-corpo, che negli ultimi decenni ha trovato ampio spazio anche nella riflessione scientifica (in particolare nelle scienze cognitive e nelle neuroscienze) nella forma del rapporto coscienza-cervello²³. La vulgata materialista o fiscalista riduce la coscienza ad un epifenomeno del cervello oppure parla di una sorta di emergenza della coscienza dalla complessificazione della struttura fisica cerebrale²⁴. Ciò ha indotto alcuni a ritenere plausibile in futuro la costruzione di un computer con un "potere causale" uguagliante o superante quello del cervello umano e la realizzazione di computer coscienti²⁵. Innanzitutto (e ciò è oramai ampiamente riconosciuto) il cervello umano, a differenza di un computer, è una struttura biologicamente complessa, non soltanto molto complicata, ovvero è una struttura olistica, dove le proprietà del tutto non sono derivabili come sommatoria di tutte le parti assemblate, ed è in simbiosi con l'ambiente circostante, una caratteristica fondamentale che permette, ad esempio, una costante rigenerazione delle cellule del corpo. Inoltre, parlare di un "potere causale" dà per scontato che sia il cervello a

²² Cfr. G. AMENDOLA, *Antropo-Logos*, cit., pp. 59-64.

²³ Cfr. D. CHALMERS, *Facing up to the problem of consciousness*, in «Journal of Consciousness Studies», 2 (1995), n. 3, pp. 200-219.

²⁴ Posizioni simili possono rinvenirsi, ad esempio, nel neuroscienziato inglese Anil Seth, che in A. SETH, *Being You. A new Science of Consciousness*, Faber & Faber Ltd, London 2021, afferma: «la mia posizione filosofica preferita, e l'assunzione predefinita di molti neuroscienziati, è il fiscalismo. Questa è l'idea che l'universo è fatto di materiale fisico e che gli stati coscienti sono identici a, o in qualche modo emergono da particolari disposizioni di questo materiale fisico» (p. 20).

²⁵ Cfr. R. KURZWEIL, *How to Create a Mind: The Secret of Human Thought Revealed*, Viking Press, New York 2012.

produrre la coscienza, mentre fenomenologicamente non dovrebbe parlarsi di causazione bensì di correlazione tra coscienza e cervello²⁶. Infine, la posizione fisicalista non offre una reale spiegazione sul perché della coscienza²⁷ e per molti studiosi rimane ancora valida la perplessità ottocentesca di Thomas Huxley secondo cui «come avvenga che qualcosa di così notevole come uno stato di coscienza sia il risultato della stimolazione del tessuto nervoso è tanto inspiegabile quanto la comparsa del Genio nella favola, quando Aladino strofina la lampada»²⁸. Non essendoci finora neppure una minima traccia su come il cervello possa generare la coscienza, vari pensatori, non solo in ambito filosofico, ma anche in ambito scientifico, stanno abbandonando l'assunto fisicalista, per riconoscere una fundamentalità o irriducibilità della coscienza stessa e, dunque, una sostanziale impossibilità nella costruzione di un computer cosciente²⁹.

Quanto abbiamo espresso finora mostra che ciò che è essenziale per l'essere umano e la sua intelligenza, come la percezione sensibile, la conoscenza, la comprensione e la coscienza, ha a che fare con qualcosa di invisibile più che di visibile, di intangibile più che di tangibile, di indescrivibile più che di descrivibile, di infinito più che di finito. D'altronde che l'essere umano sia intimamente relazionato a qualcosa di trascendente o che l'intelligenza umana abbia a che fare con una dimensione spirituale, lo sostengono da millenni le tradizioni religiose e filosofiche della Terra. Per fare qualche esempio, basti pensare al testo induista della *Kena Upanishad* risalente agli inizi del primo millennio a.C., dove questa dimensione invisibile e spirituale dell'umano è espressamente connessa allo Spirito/*Brahman*:

«Quello che è inesprimibile per la parola,
quello per mezzo del quale la parola è espressa [...].
Quello che pensa non mediante la mente,
quello dal quale la mente è pensata [...].
Quello che vede non mediante gli occhi,
quello attraverso il quale l'occhio vede le immagini [...].

²⁶ Cfr. E. KANDEL, *In Search of Memory: The Emergence of a New Science of Mind*, W.W. Norton & Company, New York 2007.

²⁷ Cfr. T. NAGEL, *Mind & Cosmos. Why the Materialist Neo-Darwinian Conception of Nature Is Almost Certainly False*, Oxford University Press, Oxford 2012.

²⁸ T.H. HUXLEY, *Lessons in Elementary Physiology*, Macmillan, London 1866.

²⁹ Cfr. F. FAGGIN, *Irriducibile. La coscienza, la vita, i computer e la nostra natura*, Mondadori, Milano 2022.

Quello che sente non mediante l'orecchio,
quello attraverso il quale l'orecchio sente ciò che ode,
Quello è il Brahman e non questo che gli uomini venerano qui»³⁰.

Anche il *Nous poietikos* di Aristotele può certamente essere interpretato in questa chiave. Infatti, scrive il filosofo greco, «non è che questo intelletto talora pensi e talora non pensi. Quando è separato, è soltanto quello che è veramente, e questo solo è immortale ed eterno (ma non ricordiamo, perché questo intelletto è impassibile, mentre l'intelletto passivo è corruttibile), e senza questo non c'è nulla che pensi»³¹. E, ancora, nella tradizione cristiana, Paolo di Tarso presenta quella tripartizione dell'umano in *pneuma* (spirito), *psyché* (anima) e *soma* (corpo) (1Ts 5,23), introducendo nella classica dualità anima-corpo lo *pneuma* come principio spirituale e divino³². Ed è proprio lo *pneuma* che viene riconosciuto come fonte della vera libertà dell'umano, perché «Il Signore è lo Spirito [*pneuma*] e, dove c'è lo Spirito [*pneuma*] del Signore, c'è libertà» (2Cor 3,17). Invece la *sarx* (carne), che si oppone allo spirito, è schiava della legge del peccato e vincola l'umano, imprigionato nei condizionamenti di questo mondo, ad agire in modo programmato e meccanico. Da qui l'invito: «Camminate secondo lo spirito [*pneuma*] e non sarete condotti a compiere i desideri della carne [*sarx*]» (Rm 5,16), perché «se vi lasciate guidare dallo Spirito [*pneuma*], non siete sotto la Legge» (Gal 5,18).

4. L'intelligenza artificiale segno di nuovi tempi e di una nuova umanità

A questo punto, all'interno degli scenari alienanti descritti nel secondo paragrafo, sembra emergere uno spazio insopprimibile di autentica libertà e, dunque, di possibilità di sfuggire alle logiche automatiche e meccaniche tipiche della struttura psico-somatica umana e agli ulteriori condizionamenti algocratici. Riteniamo inoltre che proprio la possibilità negativa di questo radicale stato di schiavitù, possa fungere da fattore di risveglio

³⁰ Kena Upanishad I, 5-8.

³¹ ARISTOTELE, *De anima*, 430 a 11.

³² A riguardo Gianfranco Ravasi sottolinea che «lo *pneuma* non è tanto il principio della vita psicofisica, ma è lo spirito divino che si effonde nella persona rendendola figlia adottiva di Dio (Rm 8,16)» (G. RAVASI, *Su anima e corpo san Paolo batte i greci*, in «Avvenire», 20 gennaio 2014).

per un'umanità ancora più libera rispetto a quanto sperimentato finora. D'altronde, come intuito dal grande poeta tedesco Hölderlin nell'inno *Patmos*, «*Wo aber Gefahr ist, wächst das Rettende auch*» («Ma dove c'è il pericolo cresce anche ciò che salva»). Ci sembra dunque sensato parlare del fenomeno dell'intelligenza artificiale in senso apocalittico/rivelativo, come uno straordinario segno dei tempi, che indica all'umano la direzione verso cui tendere in questa crisi (di crescita³³) antropologica.

In altri termini non c'è bisogno di aspettare una qualche singolarità tecnologica³⁴ per capire che il tempo che stiamo attraversando è apocalittico, in quanto già ora l'ulteriore accelerazione impressa dagli sviluppi tecnologici ci sta rivelando, *sub contraria specie*, chi siamo noi esseri umani e cosa rende una vita umana sensata sulla Terra. Non mettersi in ascolto di questo significa continuare un percorso di alienazione dalla nostra vera umanità e, dunque, tentare un sempre più improbabile equilibrio psicofisico per non sprofondare negli abissi della depressione e della disperazione. Il divario è sempre più allargato e il filo dell'equilibrista è sempre più sottile. Non ci sono più alternative tra l'incubo del dormiente e la ripresa vigile della propria umanità. Sta sempre più svanendo il limbo di cullanti e illusori sogni di assopiti sonnecchianti, mentre riecheggiano le parole apocalittiche della *Lettera ai Romani*: «è ormai tempo di svegliarvi dal sonno, perché adesso la nostra salvezza è più vicina di quando diventammo credenti. La notte è avanzata, il giorno è vicino. Perciò gettiamo via le opere delle tenebre e indossiamo le armi della luce» (*Rm* 13,11-12).

Dall'inizio delle scienze moderne in poi il cosmo è apparso sempre più come un processo evolutivo guidato da ciò che Teilhard de Chardin indicava come legge di complessificazione-coscientizzazione³⁵. Si tratta di un'evoluzione a più livelli ed integrata: cosmica, biologica, culturale, che, incrementando la complessità fisica, conduce parimenti ad un accrescimento di interiorità spirituale. L'essere umano, nell'infinitamente complesso, appare come vertice di tale processo evolutivo cosmico, mostrandosi come quella parte di universo che prende coscienza di se stesso e del suo intero processo. Tuttavia, l'umanità permane in una situazione ambigua. Come abbiamo visto, da un lato, continua a mostrare comportamenti condizionati

³³ Cfr. *Gaudium et Spes* 4.

³⁴ Cfr. R. KURZWEIL, *La singolarità è vicina*, Apogeo, Milano 2008.

³⁵ Cfr. P. TEILHARD DE CHARDIN, *Il Fenomeno Umano*, Queriniana, Brescia 2014; *Le singolarità della specie umana*, Jaca Book, Milano 2013.

da fattori ambientali, sociali e culturali e, dall'altro lato, si rivela come mai totalmente condizionabile e potenzialmente libera.

Gli sviluppi dell'intelligenza artificiale possono allora essere concepiti come un'estrema accelerazione impressa al processo evolutivo antropo-cosmico, che spinge sempre più chiaramente l'essere umano dinanzi al bivio tra un paventato assoggettamento ai sempre più pervasivi sistemi di intelligenza artificiale e l'emersione e liberazione di quel potenziale spirituale in parte assopito. Inoltre, lo sviluppo dei sistemi di intelligenza artificiale, sostituendo sempre più l'essere umano in varie attività, spinge ad una più profonda comprensione personale sull'essenza dell'essere umano e sul significato e valore del suo essere nel mondo, orientando potenzialmente verso una maggiore autenticità (in senso heideggeriano) come unica modalità di reale esistenza dell'umano. Infatti, all'interno di una concezione riduzionistico-algoritmica dell'essere umano, l'alternativa non sarebbe altro che una sopravvivenza illusoria dell'umano come macchina calcolante obsoleta, superabile da una qualche intelligenza artificiale generale del futuro, che ne orienterebbe i desideri e le azioni, ribaltando il rapporto tra artefatto e artefice, tra creatura e creatore.

In conclusione, l'intelligenza artificiale mostra così un carattere *apocalittico*, non nel senso degli immaginari scenari catastrofici post-umani tipici di tanta letteratura e filmografia fantascientifica, ma nella sua portata propriamente *rivelativa* (*apokalyptein*). Infatti, questa accelerazione evolutiva, che sta ponendo da diversi decenni in crisi tutti i sistemi consolidati in passato, spinge verso una rivoluzione antropologica e culturale, che manifesta tutti i segni del passaggio descritto nella teologia paolina da un'umanità schiava del peccato e della legge, segnata dalle opere della carne (*sarx*), ad un'umanità più cristificata, più libera, manifestante le opere dello Spirito (*pneuma*).

In tal modo, gli sviluppi dell'intelligenza artificiale possono essere letti come *segni dei tempi*, frutto di quella creatività umana e divina al tempo stesso, rivelante l'agire dello Spirito nel mondo. Un agire liberante e salvifico, che tuttavia acutizzerà drammaticamente, le conseguenze di un mondo ferito dal peccato, ambiguità tipica dei tempi apocalittici (cfr. Lc 21,10-28). Si palesa così dinanzi un bivio che richiede una svolta antropologica per evitare quell'ulteriore alienazione dell'umano e, perfino, la sua autodistruzione nel continuare a perseguire le (pseudo)logiche dello sfruttamento e del dominio delle risorse naturali (pericolo climatico), le (pseudo)logiche

neoliberiste del profitto e del dominio dell'uomo sull'uomo (pericolo economico e democratico) e le (pseudo)logiche belliche di una illimitata difesa armante per la paura dell'alterità (pericolo di una guerra nucleare), che non sono altro che le false logiche dell'uomo vecchio portate alle sue estreme e planetarie conseguenze.

La nuova epoca verso cui dovremo invece approdare riguarda primariamente un nuovo modo di essere umani nel mondo, superando questi vecchi schemi e riscoprendo innanzitutto proprio il nostro essere spirito come autocoscienza cosmica³⁶ e, dunque, quella consapevolezza di sé che agisce in ogni essere umano come crescente impulso vitale verso una libertà sempre meno condizionata dalle logiche di “questo mondo”.

Alla luce di ciò, dovrebbe essere chiaro che il comune invito delle religioni abramitiche rivolto alle imprese, agli sviluppatori, ai governanti, alla società civile e agli scienziati, per «sostenere i principi algoritmici, incorporandoli nelle politiche aziendali e sviluppando una regolamentazione internazionale per mantenerli», per «astenersi dallo sviluppo di tecnologie di intelligenza artificiale per guadagno di parte, nazionale o individuale» e per astenersi «dall'armare le tecnologie dell'IA, che minaccia la coesistenza pacifica di tutti noi»³⁷, affinché trovi una qualche concreta attuazione, richiede imprescindibilmente di intraprendere un reale cammino personale e sociale di effettiva liberazione dalle logiche dell'uomo vecchio, senza il quale è impossibile avviare quella tanto invocata quanto necessaria rivoluzione culturale e politica.

³⁶ Una tale accezione di spirito umano è presente, ad esempio, nell'*Introduzione al Cristianesimo* di Joseph Ratzinger: «esso [lo spirito] va descritto come l'essere che comprende se stesso, come essere che è presso di sé» (J. RATZINGER, *Introduzione al Cristianesimo. Lezioni sul simbolo apostolico*, Queriniana, Brescia 2010¹⁶, p. 146).

³⁷ *Joint Declaration on AI Ethics: An Abrahamic Commitment to the Rome Call*, Vatican City, 10 gennaio 2023 (la traduzione è nostra). La dichiarazione congiunta è consultabile su internet al seguente link <https://www.romecall.org/wp-content/uploads/2023/01/AI-Joint-Declaration-5-Jan1-1.pdf>.