

PHILOSOPHY

Gregorio Piaia, Roberto Prete
e Lucia Stefanutti (curr.)

INTELLIGENZA ARTIFICIALE
SVILUPPI FUTURI
E TUTELA DELLA PERSONA

Atti del Convegno di studio
(Treviso, 11 ottobre 2024)



TRIVENETO
THEOLOGY
PRESS

Questa pubblicazione afferisce alle seguenti aree dell'Anvur
(Agenzia nazionale di valutazione della ricerca universitaria):

Area 10 (Scienze dell'antichità, filologico-letterarie e storico-artistiche)

Area 11 (Scienze storiche, filosofiche e pedagogiche)

Area 12 (Scienze giuridiche)

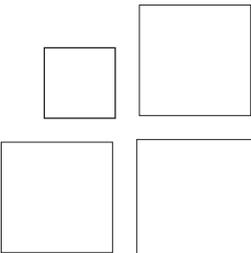
Area 14 (Scienze politiche e sociali)

PUBBLICAZIONE OPEN ACCESS
www.fttr.it - Sezione "Digital Press"

ISBN 979-12-81328-12-9

Editing: Facoltà Teologica del Triveneto
editoria@fttr.it

© 2024 Triveneto Theology Press
Facoltà Teologica del Triveneto - Padova
www.fttr.it



Gregorio Piaia, Roberto Prete
e Lucia Stefanutti (curr.)

INTELLIGENZA ARTIFICIALE SVILUPPI FUTURI E TUTELA DELLA PERSONA

Atti del Convegno di studio
(Treviso, 11 ottobre 2024)

Progetto organizzato dalla Fondazione Luigi Stefanini, realizzato con il contributo del
Ministero della Cultura - Direzione generale Educazione, ricerca e istituzioni culturali

Fondazione Luigi Stefanini



DIREZIONE GENERALE
EDUCAZIONE,
RICERCA E
ISTITUTI CULTURALI

Fondazione Luigi Stefanini

Luigi Stefanini (Treviso 1891- Padova 1956), illustre docente dell'Università patavina, tra i maggiori interpreti di Platone, fu il teorico dell'"imuginismo filosofico", e soprattutto uno degli esponenti di spicco in campo europeo dello spiritualismo cristiano e del movimento personalistico, in diretto confronto con le correnti filosofiche della prima metà del secolo XX. Studioso di estetica e importante pedagogista (in questa veste, collaboratore del Ministero della Pubblica Istruzione), si occupò di molteplici problematiche filosofiche tanto di ordine teoretico quanto di ordine storico, con una competenza e una intelligenza non comuni, e una produzione di scritti notevole.

La Fondazione Luigi Stefanini nasce nel 1996 a Treviso con lo scopo di divulgare il pensiero del filosofo e di mantenerne vivo l'interesse, rendendone accessibili le opere, specie quelle difficilmente reperibili in commercio. La Fondazione custodisce nella sua sede presso il Seminario Diocesano di Treviso (piazzetta Benedetto XI, 2) i circa tremila volumi della biblioteca personale del filosofo, nonché il suo archivio, da cui è stato tratto il volume biografico di Glori Cappello, *Luigi Stefanini, dalle opere e dal carteggio del suo archivio* (2006). In questi tre decenni di attività la Fondazione ha organizzato numerosi convegni, pubblicandone gli atti. Ne ricordiamo alcuni: *Rosmini e Stefanini* (1997), *Evoluzione e prospettive della Persona nella cultura veneta del 20° secolo* (2000), *Edith Stein e Luigi Stefanini* (2002), *Arte e Linguaggio in Luigi Stefanini* (2006). Per il settantesimo anniversario della morte del filosofo si prevede di stampare un volume di *Pagine scelte di L. Stefanini*, con il commento di una ventina di studiosi di varie università, avente lo scopo di mettere a confronto il pensiero dell'Autore con l'attualità. È prevista anche una digitalizzazione delle sue opere, che grazie alla modalità *open access* potrà allargare la platea di lettori e studiosi. Chi fosse interessato alla conoscenza del filosofo trevigiano, alla letteratura critica su di lui, ad alcune delle opere sopra citate o alle nostre prossime iniziative culturali, può consultare il sito www.fondazionestefanini.it o inviare una mail all'indirizzo fondazionestefanini@libero.it.

PHILOSOPHY - 2

RELIGIOUS STUDIES

ABSTRACT

La consapevolezza che l'Intelligenza Artificiale sia un tema che coinvolge tutti e che inciderà a fondo nella nostra vita individuale e sociale (e non solo economica) si sta sempre più diffondendo, con reazioni molto diverse che vanno dalla paura all'entusiasmo. Di qui l'opportunità di riprendere e approfondire un'analisi già avviata nel convegno *Sviluppo e diffusione dell'intelligenza artificiale e tutela della persona umana: implicazioni etico-giuridiche*, svoltosi a Treviso nell'ottobre 2023. A distanza di un anno il tema è stato affrontato in maniera ancora più incisiva, ponendo in primo piano alcune questioni di fondo che non possiamo eludere: che rapporto c'è fra la "persona" e la macchina digitale? Nel caso del soggetto umano si può ancora parlare di "anima" e "corpo", oppure vi è solo il corpo, anzi un "corpo-macchina"? Ma che significa oggi "corpo"? E la macchina è lo "specchio dell'umano" o è un pericolo mortale per la persona? La macchina più evoluta, nel suo interagire con l'uomo, può diventare un soggetto etico-giuridico? Come far fronte alla "opacità" e ai conseguenti rischi dell'IA? E che fine faranno in una società digitalizzata la democrazia politica e la legalità giuridica, così come sono state sinora intese e praticate?

Parole chiave: intelligenza artificiale, corporeità, persona, potere, cyborg, libertà, interazione uomo-macchina, tecnodiritto, etica.

The awareness that Artificial Intelligence is a topic that involves everyone and that will have a profound impact on our individual and social life (and not only economic) is increasingly spreading, with very different reactions ranging from fear to enthusiasm. Hence the opportunity to resume and deepen an analysis already started in the conference Development and dissemination of artificial intelligence and protection of the human person: ethical-legal implications, held in Treviso in October 2023. A year later, the theme has been addressed in an even more incisive way, bringing to the fore some basic questions that we cannot avoid: what is the relationship between the "person" and the digital camera? In the case of the human subject, can we still speak of "soul" and "body", or is there only the body, or rather a "body-machine"? But what does "body" mean today? And is the machine the "mirror of the human" or is it a mortal danger to the person? Can the most advanced machine, in its interaction with man, become an ethical-legal subject? How to cope with the "opacity" and consequent risks of AI? And what will happen to political democracy and legal legality, as they have been understood and practiced so far, in a digitized society?

Keywords: Artificial Intelligence, Corporeality, Person, Power, Cyborg, Man-Machine Interaction, Techno-law, Ethics.



FACOLTÀ
TEOLOGICA
DEL TRIVENETO

Questa pubblicazione afferisce all'attività di didattica e di ricerca della Facoltà Teologica del Triveneto (Padova) e di tutti gli Istituti a essa collegati (Ita e Issr)
www.fttr.it

Comitato Editoriale della Facoltà Teologica del Triveneto (Fttr) / *Editorial Board*
editoria@fttr.it

Gruppo Redazionale

Gaburro Sergio (*Ita di Verona*) – responsabile sezione Didaché
Curzel Chiara (*Issr di Trento*) – responsabile sezione Episteme
Osto Giulio (*Issr di Padova*) – responsabile sezione Praxis

Membri

Barcaro Marco (*sede Fttr*)
Boscolo Gastone (*sede Fttr*)
Dal Pozzolo Alessio (*Issr di Vicenza*)
Didonè Stefano (*direttore "Studia patavina"*)
Frausin Sergio (*Issr di Gorizia, Trieste, Udine*)
Merlo Luca (*Ita di Verona*)
Girolami Maurizio (*preside Fttr*)
Zambon Gaudenzio (*segretario generale Fttr*)
Zonato Simone (*Issr di Vicenza*)

Segreteria: Paola Zampieri (*ufficio stampa Fttr*)

Le opere proposte o richieste per la pubblicazione sono sottoposte a peer review.

SOMMARIO

<i>Gregorio Piaia</i> Prefazione	11
<i>Roberto Prete</i> Introduzione	15
<i>Paolo Moro</i> Intelligenza artificiale e libertà	19
<i>Salvatore Amato</i> L'opacità dell'intelligenza artificiale. Il problema del controllo umano significativo	33
<i>Alberto M. Gambino</i> Legalità artificiale	49
<i>Andrea Galluzzi</i> Intelligenza artificiale come specchio dell'umano. Tra consapevolezza di sé e distorsioni morfologiche	53
<i>Markus Krienke</i> Esperienza incarnata e mediazione tecnologica. Le trasformazioni digitali dell'antropologia	69
<i>Luigi Vero Tarca</i> Intelligenza artificiale o vita artificiale? Gli organismi spirituali nella nostra epoca	93

Sommario

<i>Leopoldo Sandonà</i> Per una intelligenza artificiale relazionale. Accompagnamento etico dell'evoluzione tecnologica	119
<i>Francesco Totaro</i> Noi e l'intelligenza artificiale. Dal <i>pathos</i> della paura alla confidenza critica	133
<i>Flavia Silli</i> <i>Visual culture</i> e logofobia nell'era digitale	143
<i>Tommaso Valentini</i> La democrazia nell'età della rivoluzione digitale e dell' <i>Artificial Intelligence</i>	149
<i>Gregorio Piaia</i> Note conclusive	173
Profili degli autori	179

PREFAZIONE

Gregorio Piaia

A un anno di distanza dal convegno svoltosi in questa stessa sede¹, riprendiamo l'analisi e la discussione su un tema che per la varietà e complessità dei suoi aspetti si presta a giudizi assai difformi e rischia di lasciare il comune cittadino nella totale incertezza, al punto di chiedersi se in un futuro ormai vicino saremo ridotti amaronette mosse dagli algoritmi o finiremo per innamorarci di una/un robot. L'incertezza, a sua volta, sembra cedere il passo allo sconforto: «Insomma, non ci capisco più niente con tutte queste campane che suonano in maniera contrastante» ... Confesso che anch'io di fronte alla profluvie di studi specialistici, di convegni e di interventi sulla stampa, sulla TV e sul web ho provato un po' di smarrimento, che ho superato leggendo l'introduzione di Francesco Totaro al convegno *L'antropologia dell'artificiale. Prospettive di ricerca per una ricomprensione della tecno-umanità. Configurazioni e rimodulazioni del Pathos* (svoltosi a Roma dal 19 al 21 settembre scorso) e puntando l'attenzione su due libri apparsi di recente, opera di due connazionali che hanno una profonda esperienza di studio e di insegnamento all'estero: il primo, *L'intelligenza del futuro. Perché gli algoritmi non ci sostituiranno* (Mondadori, Milano 2024), è di Paolo Legrenzi, professore emerito di psicologia all'Università Ca' Foscari di Venezia, noto specialista di scienze cognitive, formatosi a Padova e poi all'University College di Londra; l'altro volume, *Human Freedom in the Age of AI* (Routledge, New York - London 2024), è di Filippo Santoni

¹ G. PIAIA-R. PRETE-L. STEFANUTTI (a cura), *Intelligenza artificiale e tutela della persona umana. Implicazioni etico-giuridiche*, Atti del Convegno di studio (Treviso, 20 ottobre 2023), Triveneto Theology Press, Padova 2024.

de Sio, docente di Etica della Tecnologia alla Technische Universiteit di Delft in Olanda, membro di una équipe interdisciplinare promossa dalle Università di Oxford e Sidney sui problemi che lo sviluppo delle neuroscienze comportano in ordine alla responsabilità morale e giuridica.

Nella veste di coordinatore non posso qui soffermarmi a discutere su questi due libri; mi permetto solo di menzionare, a mo' di introduzione al presente convegno, le loro linee di fondo e alcuni punti significativi. Lavoro, vita quotidiana, sfera politica: sono questi – rileva Filippo Santoni de Sio – i tre ambiti in cui l'IA rappresenta una minaccia per la nostra libertà, in quanto l'IA mette a disposizione strumenti per controllare e sfruttare il lavoro (riecco gli algoritmi!), per condizionare le nostre decisioni e azioni individuali, e per compromettere la libera, effettiva partecipazione alla vita sociale e politica. In una prospettiva che si ispira alla *neo-republican tradition* ovvero al *civic republicanism*, che pone al centro del buon governo la ricerca del bene comune, la "libertà" (*freedom*) è collegata all'assenza di strutture di potere che sul piano sia personale sia collettivo potrebbero rendere i cittadini vulnerabili a una incontrollata volontà di potenza. Per raggiungere questo obiettivo è necessario che l'IA, intesa come "sistema socio-tecnologico", sia utilizzata a promuovere nei cittadini la capacità di resistere alle nuove forme di dominio che la stessa IA sarebbe in grado di realizzare o che, realisticamente, sta già ponendo in atto. Impresa tutt'altro che facile, riconosce Filippo Santoni de Sio, ma cui si può mirare da un lato favorendo la realizzazione di sistemi di IA eticamente sostenibili, dall'altro ponendo in qualche maniera sotto controllo pubblico l'applicazione dei sistemi esistenti e operanti. Essenziale diventa, a questo punto, l'aspetto politico e soprattutto giuridico, e non a caso le prime relazioni del Convegno sono svolte da studiosi del diritto. Ma essenziale è pure il grado di maturità e quindi di responsabilità (a prescindere dagli addetti ai lavori, ossia i neuroscienziati) raggiunto dal comune cittadino, cui va offerta un'adeguata informazione/sperimentazione e vanno garantiti il pluralismo dei valori e la libertà di partecipazione e di critica.

La puntuale analisi svolta da Filippo Santoni de Sio è dunque caratterizzata da quel richiamo al "dover essere" proprio di chi affronta

questioni di rilevante spessore etico. Diverso è l'approccio seguito da Paolo Legrenzi, che con il realismo dello scienziato cognitivo dissolve anzitutto l'alone fra il magico e il demoniaco che nel comune sentire avvolge l'IA. Sì, perché quella che impropriamente è chiamata Intelligenza Artificiale è senza dubbio più potente dell'intelligenza umana, ma per quanto complesso sia il suo apparato essa è priva della capacità innovativa che contraddistingue invece la "nostra" intelligenza, è priva della dimensione affettiva ed è priva di coscienza e quindi di capacità decisionale, ovvero di uscire da un programma predefinito. Ma allora, in fondo non è cambiato niente, a parte la disponibilità di macchine dalle prestazioni straordinarie? È qui il punto: queste macchine non sono solo prolungamenti o "protesi" della nostra mente e delle nostre mani, ma creano "mondi nuovi", che non abbiamo mai conosciuto nel corso della nostra lunghissima evoluzione. La distanza fra l'IA, oggetto delle scienze cognitive, e la psicologia popolare o del senso comune (quella che usiamo quotidianamente nelle nostre azioni e nelle nostre relazioni sociali) è dunque destinata ad aumentare vertiginosamente. Ed è un gap che va affrontato anzitutto sul piano formativo: posso vivere tranquillamente anche se non so nulla della fisica delle particelle, ma per la mia vita quotidiana debbo acquisire almeno alcune nozioni elementari delle scienze cognitive, giacché – per citare un aforisma che è fatto risalire a Paul Valéry (1931) – «Il futuro non è più quello di una volta»... Sul piano gestionale occorre poi rinunciare all'illusione ingenua quanto pericolosa (si pensi ai ripetuti disastri aerei causati dal Boeing 737 Max) di affidare la decisione alle sole macchine. Vi è infine, e soprattutto, il piano politico, che rinvia all'essenza del regime democratico: evitare, cioè che tecnologie ignote alla maggior parte dei cittadini siano gestite e utilizzate da persone o gruppi di fatto anonimi, il che creerebbe nuove disuguaglianze e nuove schiavitù, molto più profonde e perverse – sottolinea Paolo Legrenzi con una punta di ironia alla pagina 41 – di quelle cui si ribellò il famoso Spartaco nel I sec. a.C.: «Il fatto più grave, a differenza dei tempi di Spartaco, è che questo sfruttamento possa avvenire all'insaputa degli sfruttati che, inconsapevoli e beati, gioiscono di tale sottomissione». E questa, daccapo, è una sfida che va

affrontata dai giuristi accanto ai politici. Lascio pertanto la parola all'avv. Roberto Prete, vicepresidente della sezione trevigiana dell'Unione Giuristi Cattolici Italiani, che insieme con la Fondazione Luigi Stefanini ha organizzato questo convegno, i cui atti sono ora pubblicati grazie al contributo finanziario del Ministero della Cultura (Direzione generale Educazione, ricerca e istituti culturali).

INTRODUZIONE

Roberto Prete

Abbiamo sentito l'esigenza di un nuovo convegno su un tema di stretta attualità, di cui si avverte l'importanza e la complessità delle implicazioni. È di alcuni giorni fa la notizia che il Premio Nobel per la Fisica 2024 è stato assegnato a John Hopfield dell'Università di Princeton (Stati Uniti) e a Geoffrey Hinton dell'Università di Toronto (Canada). Il Comitato Nobel ha specificato che il premio è stato attribuito per le scoperte ed invenzioni che consentono l'apprendimento automatico tramite reti neurali artificiali. Queste reti rappresentano un modello computazionale, ispirato alla struttura del cervello umano, dove i nodi fungono da neuroni, interagendo tra loro in modo simile alle sinapsi. È quindi innegabile che l'Intelligenza Artificiale è un nuovo mezzo di progresso, con il potere di influire a fondo su diversi settori del mondo umano. Tutto questo però richiede la salvaguardia essenziale della dignità umana. Papa Francesco nel discorso tenuto il 14 giugno 2024 al Forum Intergovernativo G7 sul tema dell'Intelligenza Artificiale, tenutosi a Borgo Egnazia (Br), ha messo in rilievo questo importante valore. Riporto alcuni brani significativi del suo discorso:

«La portata di queste complesse trasformazioni è ovviamente legata al rapido sviluppo tecnologico dell'intelligenza artificiale stessa. Proprio questo vigoroso avanzamento tecnologico rende l'intelligenza artificiale *uno strumento affascinante e tremendo* al tempo stesso ed impone una riflessione all'altezza della situazione. [...] Anzi, non di rado, proprio grazie alla sua radicale libertà, l'umanità ha pervertito i fini del suo essere trasformandosi in nemica di sé stessa e del pianeta. Stessa sorte possono avere gli strumenti tecnologici. Solo se sarà garantita la loro vocazione al servizio dell'umano, gli strumenti tecnologici riveleranno non solo la grandezza e la dignità unica dell'essere umano, ma anche

il mandato che quest'ultimo ha ricevuto di «coltivare e di custodire» (cfr. *Gen 2,15*) il pianeta e tutti i suoi abitanti. [...] Di fronte ai prodigi delle macchine, che sembrano saper scegliere in maniera indipendente, dobbiamo aver chiaro che all'essere umano deve sempre rimanere la decisione, anche con i toni drammatici ed urgenti, con cui a volte questa si presenta nella nostra vita. Condanneremmo l'umanità a un futuro senza speranza, se sottraessimo alle persone la capacità di decidere su loro stesse e sulla loro vita, condannandole a dipendere dalla scelta delle macchine. Abbiamo bisogno di garantire e tutelare uno spazio di controllo significativo dell'essere umano sul processo di scelta dei programmi di Intelligenza artificiale: ne va della stessa dignità umana».

Questi temi sono stati affrontati in particolare da padre Paolo Benanti, docente alla Pontificia Università Gregoriana, membro del *New Artificial Intelligence Advisory Board*, sorto nel 2023 su iniziativa del segretario generale dell'Onu Antonio Guterres, e presidente, fra l'altro, della Commissione sull'intelligenza artificiale per l'informazione del Dipartimento per l'informazione e l'editoria della Presidenza del Consiglio dei Ministri. Nel libro *Human in the loop. Decisioni umane e intelligenze artificiali*, pubblicato nel 2022, ha rilevato che

«le intelligenze artificiali permeano molti aspetti del vivere quotidiano (per esempio per fare una ricerca in internet), mediante l'azione di vari algoritmi [...] Queste tecnologie, proprio perché presenti sul fondo dell'esistenza, diventano pressoché invisibili e ci sono sconosciute nella loro vera natura. Cercare di rendere visibile e comprensibile l'azione di questi strumenti onnipresenti e chiedersi cosa fare per gestirli e come non estromettere l'uomo dal decidere, è l'obiettivo importante. Occorre affrontare la sfida di mantenere l'umanità capace di controllo, in un'epoca in cui la macchina si fa capace di surrogare le decisioni umane».

In un altro volume, *Noi e la macchina* (2024), padre Paolo Benanti si è così espresso:

«Tutto il mondo della tecnologia viene descritto dalla categoria dell'innovazione. Ma esiste anche un'innovazione di tipo sociale, che oggi tende a sfuggirci di vista, morale ed etico, che accompagna il progresso scientifico e tecnico della nostra società. Innovazione tecnica e cambiamenti socio-culturali vanno sempre di pari passo, e bisogna tenerne

conto. Se continueremo a considerare l'innovazione solamente dalla prospettiva tecnologica, rischiamo di non riuscire più a percepire la portata in termini di trasformazione sociale, risultando incapaci di orientarne gli effetti verso il bene. [...] Parlare di sostenibilità digitale significa ribaltare la questione ed affrontarla con le categorie dell'etica, non mettere la capacità tecnica al centro dell'attenzione, bensì considerare la persona come fine ultimo, che deve orientare il progresso tecnologico».

La necessità di un controllo dello sviluppo di queste tecnologie è dunque molto sentita. Anche mons. Vincenzo Paglia, presidente della Pontificia Accademia per la Vita, nel libro *Algoritmo della vita. Etica ed intelligenza artificiale*, pubblicato quest'anno, ha fatto presente come l'Intelligenza Artificiale non possa avere un uso acritico e irrazionale: «Attualmente tutta la materia è in mano a tre aziende a livello mondiale ed è indispensabile trovare un accordo, affinché i dati rimangano in mano pubblica. Occorre un nuovo umanesimo, che valorizzi l'umano comune in tutti. Questo potrà essere acquisito in un quadro giurisdizionale internazionale». Si sottolinea, in altre parole, la necessità urgente di una seria regolamentazione a livello europeo e mondiale. In Europa esiste l'*European AI Act*, che però dà una regolamentazione in ordine solo alla tutela dei dati personali e della privacy, oltre alla salvaguardia del libero mercato. Questa normativa limitata è sentita carente da molte parti e si avverte la necessità di una regolamentazione più estesa.

Recentemente è apparso, su un'intera pagina del *Corriere della Sera* del 21 settembre 2024, un avviso a pagamento dal titolo *L'Europa ha bisogno di certezza normativa sull'Intelligenza Artificiale* e dal sottotitolo *Con una regolamentazione frammentaria l'Unione Europea rischia di non avere un ruolo attivo nell'era dell'Intelligenza Artificiale*. L'avviso è sottoscritto dai rappresentanti di una serie di aziende italiane ed europee, attive sul mercato mondiale (Pirelli, Prada Group, Ericsson, Essilor Luxottica, ICREA, etc.). Una richiesta più che legittima, naturalmente, purché l'auspicata certezza normativa, oltre a favorire lo sviluppo della tecnologia, provveda anche a tutelare i diritti di tutte le persone. Non sono infatti mancati i richiami al valore incommensurabile della persona umana. Anche il cardinale Pietro Parolin, segretario di Stato, al 52°

Congresso nazionale della Società italiana di medicina fisica e riabilitativa, tenutosi recentemente a Padova, ha affermato: «Sì all'intelligenza artificiale se aiuta nelle cure e sopperisce a deficit fisici e rende più facile la vita. Ma un robot non potrà mai sostituire la socialità». Anche questa tecnologia ha dunque dei limiti, che vanno sempre tenuti presenti. Concludo augurando a tutti di trarre un arricchimento culturale da questo convegno.

INTELLIGENZA ARTIFICIALE E LIBERTÀ

Paolo Moro

ABSTRACT

Si esamina anzitutto la possibilità di considerare il robot un agente morale e dunque centro di autodeterminazione e d'imputazione di una responsabilità etica e giuridica. Si evoca poi il conflitto fra il libero arbitrio della persona umana e l'autonomia dei sistemi di IA, in vista di un'etica della sostenibilità tecnologica che oltrepassi il soggettivismo moderno e consenta un'equilibrata convivenza con le macchine intelligenti, ma anche la possibilità del diritto di dissociarsi dal "metaverso globale".

Parole chiave: IA, Intelligenza meccanica, Libertà negativa, Libertà assoluta, Libertà umana.

First we examine the possibility of considering the robot as a moral agent and, therefore, as a center of self-determination and imputation of ethical and juridical responsibility. Secondly, conflict is evoked between the free will of the human person and the autonomy of the AI, which requires an ethics of technological sustainability that goes beyond the modern subjectivism and allows not only a balanced coexistence with intelligent machines, but the possibility of the right to dissociate from the "global metaverse".

Keywords: AI, Negative freedom, Absolute freedom, Human freedom.

1. Macchine "libere"

L'avanzamento tecnologico dell'intelligenza artificiale sta spingendo i sistemi mecatronici avanzati a trasformarsi in forme robotiche immateriali e complesse, che assomigliano a persone artificiali. Questo orientamento ha portato ad umanizzare superficialmente le macchine, attraverso un'imitazione ingenua del comportamento umano, e ha trascurato le difficoltà teorica e pratica di stabilire un accettabile paragone

tra intelligenza artificiale e coscienza umana, tuttora presupposta sulla base di una concezione riduttiva della soggettività umana. Tuttavia, è ormai assodato che le macchine intelligenti dell'era contemporanea non sono soltanto automi che sostituiscono in modo meccanico e ripetitivo l'uomo in lavori pesanti ma, grazie all'evoluzione degli studi di intelligenza artificiale, hanno sviluppato una presupposta abilità di deliberare e decidere, apparentemente simile a quella umana, adombrando la possibilità di individuare una paradossale libertà del robot e di elaborarne un'etica¹.

È sempre più evidente che i robot più raffinati, ritenuti capaci non solo di eseguire calcoli complessi ma anche di esibire condotte reattive in risposta a stimoli esterni, possono essere paragonati a soggetti morali e possono essere considerati centri di autodeterminazione volitiva e d'imputazione di responsabilità etica e giuridica, con conseguenze problematiche soprattutto nel diritto penale².

L'alleanza tecnologica tra informatica e cibernetica ha permesso di costruire sistemi che possono elaborare dati in base al modello computazionale ed esprimere una facoltà simile alla volontà attraverso la retroazione (*feedback*), ossia la capacità autonoma di adattarsi agli stimoli ambientali e di adottare comportamenti conseguenti. Il paragone tra uomo e macchina, ipotizzato dall'odierno postumanesimo³, è l'esito di una mentalità che estende tale raffronto al valore della libertà. Se dotata di facoltà soggettive analoghe o anche superiori a quelle umane, la macchina intelligente ed evoluta può diventare titolare di una indefinita autodeterminazione individuale e, quindi, di libero arbitrio, come

¹ Cf. P. MORO, *Libertà del robot? Sull'etica delle macchine intelligenti*, in R. BRIGHI-S. ZULLO (curr.), *Filosofia del diritto e nuove tecnologie. Prospettive di ricerca tra teoria e pratica*, Atti del XXIV Congresso della Società Italiana di Filosofia del Diritto, Aracne, Roma 2015, 525-544.

² S. RIONDATO, *Robot: talune implicazioni di diritto penale*, in P. MORO-C. SARRA (curr.), *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, FrancoAngeli, Milano 2017, 85-98.

³ R. PEPPERED, *The Posthuman Manifesto*, in *Kritikos* 2 (2005) 1-16.

ha sostenuto l'inventore del sintagma "intelligenza artificiale", John McCarthy⁴.

Come è noto, la caratteristica tipica del sistema di intelligenza artificiale è la sua capacità di apprendimento automatico e interattivo (*interactive machine learning*), che è il processo con il quale il robot incrementa progressivamente la propria abilità e la propria autonomia operando insieme con l'essere umano, con l'esigenza di elaborare norme giuridiche che regolino gli inevitabili conflitti generati dall'interazione tra soggetti ed entità sempre meno distinte⁵. Le macchine intelligenti ed autonome mostrano la loro presenza in misura ormai determinante in molti campi dell'esperienza sociale (come quello industriale o domestico) e invadono i pensieri e le azioni propriamente umane, il cui condizionamento tecnologico comprime quella libertà che, paradossalmente, è sempre più concessa proprio alle macchine stesse.

Si è affermato che gli studi e le applicazioni relative all'apprendimento automatico delle macchine (*machine learning*) possono consentire di raggiungere l'autentico "algoritmo dominante"⁶, sicché l'agente robotico potrebbe acquisire forme di responsabilità e di libertà, assumendo doveri e diritti soggettivi molto simili a quelli umani. Questa possibilità di attribuire una personalità giuridica ai robot particolarmente avanzati e sempre più simili agli umani è già acquisita nel dibattito scientifico, nel quale si applicano categorie che tradizionalmente identificano la tutela del soggetto umano, come i diritti soggettivi, ai sistemi di intelligenza artificiale⁷.

⁴ J. McCARTHY, *Free Will - even for Robots*, in *Journal of Experimental and Theoretical Artificial Intelligence*, 12/3 (2000) 341-352.

⁵ M. A. GOODRICH-A. C. SCHULTZ, *Human-Robot Interaction: A Survey*, in *Foundations and Trends in Human-Computer Interaction*, 1/3 (2007) 203-275.

⁶ P. DOMINGOS, *The Master Algorithm. How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, New York 2015.

⁷ B.-J. KOOPS-M. HILDEBRANDT-D.-O. JACQUET-CHIFFELLE, *Bridging the Accountability Gap: Rights for New Entities in the Information Society?*, in *Minnesota Journal of Law, Science & Technology* 11/2 (2010) 497-561; G. SARTOR, *Gli agenti software: nuovi soggetti del ciberdiritto?*, in *Contratto e impresa* 18/2 (2002) 465-499; L. LOMBARDI VALLAURI, *Neuroni, mente,*

Il Comitato Nazionale per la Bioetica e il Comitato Nazionale per la Biosicurezza, le Biotecnologie e le Scienze della Vita hanno notato che lo sviluppo di sistemi informatizzati di intelligenza artificiale, applicati a una molteplicità di ambiti sociali ed economici, rende «del tutto insufficiente applicare ai robot i regimi e le dottrine giuridiche esistenti alla luce delle tradizionali categorie giuridiche (persone fisiche, persone giuridiche, beni mobili)»⁸.

2. Macchine “sociali”

Poiché la capacità di apprendimento automatico incrementa progressivamente l'autonomia del robot anche quando coopera con l'essere umano, come avviene nella robotica collaborativa⁹, l'impatto della regolazione algoritmica della società (per esempio, con attività di *crowdsourcing* o di gestione della reputazione) trasforma il sistema di intelligenza artificiale in una “macchina sociale”, che genera evidenti conseguenze etiche e giuridiche per la tutela dell'autonomia individuale e dell'ordine pubblico¹⁰. L'elaborazione e l'applicazione di una *lex robotica* che possa conciliare potenziali conflitti intersoggettivi tra uomini e macchine diventa un'urgente problematica del tecnodiritto

anima, algoritmo: quattro ontologie, in L. LOMBARDI VALLAURI (cur.), *Logos dell'essere. Logos della norma*, Adriatica, Bari 199, 571-601; G. TADDEI ELMI, *I diritti dell'intelligenza artificiale tra soggettività e valore: fantadiritto o ius condendum?*, in L. LOMBARDI VALLAURI (cur.), *Il meritevole di tutela*, Giuffrè, Milano 1990, 685-711.

⁸ Comitato Nazionale per la Bioetica - Comitato Nazionale per la Biosicurezza, le Biotecnologie e le Scienze della Vita, *Sviluppi della robotica e della roboetica*, Parere 17 luglio 2017, 35 [http://bioetica.governo.it/media/1392/p129_sviluppi-della-robotica-e-della-roboetica-gruppo-misto_it-cnb.pdf].

⁹ M. A. GOODRICH-A. C. SCHULTZ, *Human-Robot Interaction: A Survey*, in *Foundations and Trends in Human-Computer Interaction* 1/3 (2007) 203–275.

¹⁰ N. CRISTIANINI-T. SCANTAMBURLO, *On social machines for algorithmic regulation*, in *AI & Society* 35 (2020) 645-662 [<https://doi.org/10.1007/s00146-019-00917-8>].

contemporaneo¹¹, soprattutto al fine di preservare la tutela e l'integrità delle persone attraverso l'imposizione di limiti etici e giuridici alle applicazioni della ricerca tecnologica¹². In effetti, il robot che interagisce con l'uomo si troverà di fronte ad una serie di dilemmi morali che imporranno alla macchina intelligente la necessità di deliberare una scelta fra più alternative, come, per esempio, costringere un anziano ad assumere un farmaco. L'implementazione di regole morali che permettono al robot di prendere decisioni consapevoli e comportarsi in modo giusto oppure sbagliato può essere programmata con algoritmi di istruzione ma anche di apprendimento dalla condotta che l'automa pone in essere, però tale questione potrebbe andare oltre una forma controllabile di intelligenza artificiale¹³.

Difatti, le procedure formali di apprendimento automatico restano inferenze ipotetiche e non possono interpretare o risolvere problemi complessi della vita reale, come i conflitti di valore, che manifestano costante variabilità ed incertezza e che sono difficili da codificare attraverso misurazioni o funzioni numeriche¹⁴. D'altronde gli studi sull'intelligenza artificiale tendono a ridurre la definizione della "volontà" del robot alla reazione automatica del dispositivo agli stimoli fisici che riceve dall'ambiente, aderendo alla mentalità del razionalismo matematico che, immaginando che la macchina sia un "cervello elettronico", riduce la mente umana alla sola funzione calcolante¹⁵.

Questa visione presuppone l'adesione alla tesi dell'intelligenza artificiale "forte" e propone un paragone che influisce tuttora sulle ricerche

¹¹ P. MORO-C. SARRA (curr.), *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, FrancoAngeli, Milano 2017.

¹² A. FABRIS-G. TAMBURRINI, *Robot al di qua del bene e del male*, a cura di R. PRESILLA. in *Vita e Pensiero* 3 (2006).

¹³ W. WALLACH-C. ALLEN, *Moral machines. Teaching robots right from wrong*, Oxford University Press, Oxford 2009.

¹⁴ Z. LIPTON, *The mythos of model interpretability*, in *ICML Workshop on Human Interpretability*, 2016, arXiv:1511.03677.

¹⁵ P. MORO, *Algoritmi e pensiero giuridico. Antinomie e interazioni*, in *MediaLaws. Rivista di diritto dei media* 3 (2019) 11-22.

riguardanti gli agenti artificiali intelligenti, l'interazione tra robot ed esseri umani e i sistemi multiagente, postulando l'attribuzione, quantomeno ad alcune entità di questo insieme di sistemi cibernetici, di alcune facoltà o valori tipici del soggetto umano, come la stessa libertà. È questa la concezione del razionalismo moderno, secondo cui l'uomo sarebbe capace di conoscere ciò che è bene e ciò che è male attraverso l'uso della propria ragione e che ha contribuito a fondare i programmi di intelligenza artificiale sul metodo scientifico¹⁶, riconducendo a schemi comuni di origine algoritmica le macchine e i sistemi biologici capaci di esibire un comportamento finalizzato, confermando il presupposto matematico della cibernetica¹⁷.

La costruzione del cosiddetto *imitation game* da parte di Alan Turing è indubbiamente condizionata da un'interpretazione della coscienza e della volontà della macchina basata sugli effetti della condotta reattiva agli impulsi esterni: tale interpretazione deriva dalla psicologia comportamentale ed è fondata sull'assunto che la condotta esplicita sia l'unica unità di analisi scientifica, in quanto direttamente osservabile, mentre l'introspezione non potrebbe fornire alcun dato affidabile¹⁸. Nell'intelligenza artificiale, dunque, le teorie degli agenti e delle reti neurali sembrano dipendere da questo approccio riduzionista, come l'ipotesi descrittiva della "società della mente" di Marvin Minsky, secondo il quale il cervello è una macchina altamente complessa paragonabile a una società organizzata da una rete di interconnessioni gerarchiche e competitive tra una molteplicità di componenti fortemente diversi fra loro, denominati appunto "agenti della mente"¹⁹.

¹⁶ D.A. GILLIES, *Artificial intelligence and scientific method*, Oxford University Press, Oxford, 1996, trad. it.: *Intelligenza artificiale e metodo scientifico*, Cortina, Milano 1998.

¹⁷ G. TAMBURRINI, *I matematici e le macchine intelligenti. Spiegazione e unificazione nella scienza cognitiva*, Bruno Mondadori, Milano 2002.

¹⁸ J.B. WATSON, *Psychology as a behaviorist view*, in *Psychological review* 20/2 (1913) 158-177.

¹⁹ M.L. MINSKY, *The Society of Mind*, Simon and Schuster, New York 1986.

3. Libertà assoluta

La forma con la quale gli scienziati tentano di attribuire un libero arbitrio alla "volontà" della macchina intelligente è riduttivo, giacché riguarda l'accezione puramente negativa del termine "libertà", che è stato concepito in epoca moderna principalmente come affrancamento da un potere esterno al soggetto. La libertà negativa si costituisce nell'epoca della liberazione dell'"individuo" (che identifica specificamente la persona umana nel pensiero della modernità) da ogni oppressione all'esplicarsi del proprio volere, e matura contestualmente con l'evoluzione della teoria dei diritti soggettivi e delle dottrine dei riconosciuti maestri del Giusnaturalismo del Seicento e del Settecento (Hobbes, Locke e Rousseau). La prospettiva individualistica della libertà negativa, che nella coscienza comune irriflessa si identifica con il diritto di fare o di omettere ciò che si vuole, si diffonde nella cultura giuridica e politica europea del diciottesimo secolo e diventa il supporto intellettuale dell'ideologia dei diritti nella Rivoluzione francese²⁰. È emblematico il riassunto di questa concezione fatto nella nota conferenza pronunciata all'*Athénée Royal* di Parigi nel 1819 da Benjamin Constant, il quale promuove l'idea di libertà come autodeterminazione rilevando che «la nostra libertà deve fondarsi sul pacifico godimento dell'indipendenza privata»²¹.

La possibilità di compiere atti incondizionati di autodeterminazione, che presuppone l'assenza di vincoli alla capacità di deliberare e di scegliere, accentua il profilo psicologico, in cui l'idea della libertà morale appartiene all'individuo, considerato razionalmente capace di scegliere il proprio bene e i mezzi per conseguirlo attraverso la piena autarchia della volontà. L'art. 13 della Costituzione italiana accoglie questa libertà negativa ove inaugura la disciplina dei rapporti civili e politici della

²⁰ A. MARTINELLI-S. VECA-M. SALVATI, *Progetto 89. Tre saggi su libertà, eguaglianza, fraternità*, Il Saggiatore, Milano 2009.

²¹ B. CONSTANT, *De la liberté des Anciens comparée à celle des Modernes*, Paris 1819, trad. it.: *Discorso sulla libertà degli antichi paragonata a quella dei moderni*, Editori Riuniti, Roma 1992, 15-16.

Carta repubblicana, affermando nettamente nel primo comma che «la libertà personale è inviolabile», alludendo così all'aggettivo già significativamente utilizzato nell'art. 2 per designare i diritti umani. Infatti, riferendosi alla libertà fisica, il secondo comma dell'art. 13 spiega che «non è ammessa forma alcuna di detenzione, di ispezione o perquisizione personale, né qualsiasi altra restrizione della libertà personale, se non per atto motivato dell'Autorità giudiziaria e nei soli casi e modi previsti dalla legge».

Questa prospettiva qualifica la libertà soggettiva come affrancamento da una costrizione esteriore ed è il presupposto antropologico del tentativo, che si sta manifestando nell'epoca contemporanea, di definire la soggettività robotica. È la concezione razionalista della libertà, fondata sul paradigma teorico dell'assolutezza. Nel senso più generico, "assoluto" (*ab solutus*) è ciò che è sciolto da vincoli ed è indipendente da ostacoli esterni: con l'affermarsi dell'idea, molto diffusa nel Cinquecento e nel Seicento, che l'etica e la politica possano essere spiegate con i procedimenti della conoscenza matematica e fisica della realtà, la qualificazione generica dell'assolutezza diviene tipica di ogni atto individuale che, così, è pensato come incondizionato.

Questa visione nasce dalla convinzione, ampiamente diffusa in età moderna, che la realtà sia dominabile dall'uomo, il quale sarebbe in grado di analizzarla per scomporla e ricomporla con metodo matematico al fine di apprenderne la struttura e di intervenire operativamente su di essa. Si costituisce così l'idea che il soggetto disponga di una libertà creatrice e autosufficiente che trova il proprio realizzarsi nel rapporto di supremazia sulla natura, oggettivamente concepita come un meccanismo inanimato e compiutamente controllabile dalle facoltà soggettive. Questa potenza trasformatrice dell'individuo, che è ritenuto in grado di dominare gli oggetti della sua conoscenza e della sua volontà, risale alla logica possessiva della modernità²² e alla dimensione antropologica dell'*homo faber fortunae suae*, secondo cui il soggetto è un atomo autosufficiente che riconosce nella propria e nell'altrui

²² P. BARCELLONA, *L'individualismo proprietario*, Bollati Boringhieri, Torino 1987, 111-112.

soggettività un oggetto analizzabile, dominabile e manipolabile²³.

Nella cultura giuridica e politica dell'età della rivoluzione scientifica germina l'idea che l'individuo, attraverso la costruzione della struttura dello Stato, possa darsi una legge positiva che regoli i suoi comportamenti nella società indipendentemente da qualsiasi condizionamento a lui estraneo, giacché la ragione appare in grado di ritrovarne la matrice nell'ordine della natura investigata con criteri matematici e geometrici²⁴. In questo contesto storico e filosofico il paradigma della libertà assoluta è confezionato dall'empirista inglese Thomas Hobbes, secondo il quale la filosofia politica è una scienza civile²⁵ fondata sulla matematica, giacché l'uomo può riconoscere e produrre le regole più utili per sé e per gli altri al fine operativo di giustificare la presenza dello Stato. Nel *De Cive* si legge che «volendo definire la libertà, si deve dire che essa non è altro che l'assenza di impedimenti ai nostri movimenti»²⁶ e, similmente, nel *Leviatano* si afferma che «secondo il significato proprio del termine, si intende per libertà l'assenza di impedimenti esterni (*absence of external impediments*)»²⁷.

È questa la condizione naturale dell'uomo, che non può incontrare intralci alla propria condotta perché la volontà è sempre istintivamente diretta a proteggere la propria integrità fisica, sicché la ragione deve adeguare il proprio giudizio a questo coerente principio regolativo di un'esistenza senza ostacoli concreti. Pertanto, si può notare che l'imitazione delle principali qualità psichiche da parte delle macchine, guidata da scienziati e programmatori sicuramente influenzati dal pensare moderno, si sviluppa in base al preconconcetto della libertà assoluta ed appare l'esito narcisistico di una prospettiva antropocentrica e riduzionistica del soggetto umano, che ricompone le proprie funzioni con la

²³ F. ZANUSO, *Conflitto e controllo sociale nel pensiero giuridico-politico moderno*, Cleup, Padova 1993, 16 ss.

²⁴ C. TAYLOR, *Il disagio della modernità*, Laterza, Roma-Bari 1994.

²⁵ Cf. T. SORELL, *Hobbes*, Routledge, London-New York 1991.

²⁶ T. HOBBS, *De Cive*, IX, trad. di N. Bobbio, UTET, Torino 1948, 219.

²⁷ T. HOBBS, *Leviatano*, XIV, trad. di M.V. Predaval, Laterza, Roma-Bari 1996, 105.

fabbricazione di frammenti digitali del soggetto²⁸. In effetti la coscienza non è riducibile a una sola delle sue funzioni e, dunque, il tentativo di paragonare la mente umana all'intelligenza artificiale è destinato a fallire: la coscienza è una proprietà dell'intera persona e deve essere considerata nella totalità che contraddistingue l'intelligenza umana, che non può essere riprodotta, quantomeno nella sua completezza, attraverso modelli computazionali anche molto evoluti, giacché essa genera un problema che non è computabile²⁹.

4. Libertà futura

La società futura annuncia il sorgere di nuove relazioni e nuovi conflitti tra uomini e macchine intelligenti. La "personalità elettronica" dei robot potrebbe impattare sulla vulnerabilità degli esseri umani, che insegnano l'autonomia alle macchine ma conservano la propria libertà³⁰, suggerendo al legislatore o al giudice di trovare un giusto temperamento degli interessi nel rispetto del principio *neminem laedere* e del valore della dignità individuale. Dunque, la risemantizzazione dei fenomeni che riguardano l'intelligenza artificiale procede da un adeguato ripensamento della nozione della libertà umana, che il pensiero moderno ha ridotto alla sua dimensione negativa e assoluta: il libero arbitrio non è solo un fatto biologico, definito dalla reazione automatica ed estrinseca agli stimoli fisici ricevuti dall'ambiente esterno, ma si presenta come atto soggettivo di introspezione cosciente e relazionale.

Forse la macchina può pensare, ma in base al codice binario, mentre l'uomo è dotato di una mente colorata, come ricorda Omero a proposito

²⁸ P. MORO, *Persona elettronica. Una finzione giuridica per l'intelligenza artificiale*, in *L'Ircocervo*, 1/2024, 372-389.

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

³⁰ S. AMATO, *Biodiritto 4.0. Intelligenza artificiale e nuove tecnologie*, Giappichelli, Torino 2020, 112.

del multiforme ingegno di Ulisse³¹, e vive di plurime dimensioni della soggettività che, come ha accertato la psicoanalisi del Novecento, provengono anche dall'inconscio. Secondo Sigmund Freud, la vita psichica si svolge in modalità performative anche nelle sue forme secondarie – *lapses*, dimenticanze, sogni – che malcelano un'organizzazione alternativa della razionalità³²: tali forme del pensare sono del tutto ignote all'intelligenza meccanica che, quando sbaglia, esegue un'operazione che è giudicata falsa in base alla logica di Boole e, dunque, non commette affatto un *lapses*. Nel suo famoso esperimento mentale noto come "stanza cinese" ed illustrato per criticare i sostenitori dell'intelligenza artificiale "forte", John Searle ha contestato che la mente possa essere un programma, perché il *computer* non capisce ciò che sta facendo quando elabora un'operazione mentale attraverso il linguaggio simbolico, essendovi una distanza incolmabile tra sintassi e semantica, ossia tra struttura grammaticale del linguaggio e comprensione significativa delle operazioni combinatorie³³.

La possibilità tecnologica delle macchine intelligenti di rendersi autonome e di prendere decisioni anche errate o ingiuste non è una questione computabile, ma sta portando alla costituzione di una vera e propria "etica dei robot", alla cui azione saranno attribuiti doveri e forse anche diritti. Senza una riflessione critica sulla libertà umana, l'agente intelligente potrà distinguere non solo il vero dal falso, ma anche il giusto dall'ingiusto, e forse giungerà il tempo della soggettività robotica³⁴. Negli ultimi anni il progresso dei sistemi di intelligenza artificiale ha suggerito una riflessione ampia e articolata sui limiti etici delle applicazioni di questa tecnologia, che rischiano di pregiudicare la dignità umana e i diritti fondamentali. È stato ritenuto che la questione etica possa

³¹ P. CITATI, *La mente colorata. Ulisse e l'Odissea*, Mondadori, Milano 2002.

³² S. FREUD, *Vorlesungen zur Einführung in die Psychoanalyse* (1915), trad. it.: *Introduzione alla psicoanalisi*, in *Opere di Sigmund Freud*, VIII, Bollati Boringhieri, Torino 2000.

³³ J.R. SEARLE, *Minds, brains, and programs*, in *Behavioral and Brain Sciences* 3/3 (1980) 417-457.

³⁴ P. MORO, *Macchine come noi. Natura e limiti della soggettività robotica*, in U. RUFFOLO (cur.), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Giuffrè, Milano 2020, 45-62.

risolversi con l'applicazione di un principio giustificativo di comprensibilità (*explicability*), inteso non solo in senso epistemologico come "intelligibilità" (risposta alla domanda "come funziona l'intelligenza artificiale?") ma anche in senso morale di "responsabilità" (risposta alla domanda "chi è responsabile del modo in cui funziona l'intelligenza artificiale?")³⁵.

Il potenziamento sempre più accelerato dell'intelligenza artificiale e delle sue applicazioni sta generando grandi benefici di immenso impatto sociale ed economico, ma sta impoverendo la libertà e altre dimensioni relazionali della vita umana, sempre più oggettivata e priva di comunicazione, dipendendo dal presupposto individualista del razionalismo cibernetico. È il tempo di un ripensamento critico di questa visione antropologica, elaborata dal grande progetto computazionale della modernità, per tentare di risolvere con metodi alternativi e modelli innovativi le difficoltà poste dalle applicazioni della cibernetica e dell'intelligenza artificiale. Prima d'altro, appare necessario condurre uno sviluppo controllato delle capacità umane nel robot non tanto attraverso il potenziamento (*enhancement*) delle facoltà umane, ma con la realizzazione (*achievement*) delle varie qualità della persona, tenendo conto che, come sembra inconfutabile, la soggettività non è uno *standard* misurabile né un programma determinabile da algoritmi.

Il conflitto tra la "libertà" della comunità umana e l'"autonomia" dei sistemi di intelligenza artificiale, immaginato già nel secolo scorso dalla più raffinata letteratura fantascientifica³⁶, impone di riconsiderare un'"etica della sostenibilità tecnologica" che permetta un superamento critico dell'individualismo moderno e favorisca una convivenza equilibrata e responsabile con le macchine intelligenti³⁷. Pertanto, nel rispetto della

³⁵ L. FLORIDI-J. COWLS, *A Unified Framework of Five Principles for AI in Society*, in *Harvard Data Science Review* 1/1 (2019) 2-15.

³⁶ P.K. DICK, *Do Androids Dream of Electric Sheep?*, Doubleday and Company, New York 1968.

³⁷ P. MORO, *Alle frontiere della soggettività: indizi di responsabilità delle macchine intelligenti*, in U. RUFFOLO (cur.), *XXVI Lezioni di Diritto dell'Intelligenza Artificiale*, Giappichelli, Torino 2021, 37-47.

dignità intangibile della persona, la libertà giuridica rimane sempre relazionale e profondamente umana. È indubbio il diritto di accesso alle tecnologie emergenti, come l'intelligenza artificiale, ma è anche inviolabile il diritto di dissociarsi dalla trasformazione digitale e di restare fuori dal "metaverso globale".

L'OPACITÀ DELL'INTELLIGENZA ARTIFICIALE. IL PROBLEMA DEL CONTROLLO UMANO SIGNIFICATIVO

Salvatore Amato

ABSTRACT

I modelli di apprendimento profondo generativo impongono una nuova visione dell'IA in relazione all'autonomia, alle prestazioni e alla capacità di autotrasformazione. L'enorme assimilazione di dati e la continua capacità di rielaborazione hanno complessificato la complessità. Ogni dispositivo ne sviluppa altri e poi altri ancora fino a una concatenazione strutturale che tende a diventare un mondo a sé. È quindi sempre più difficile rompere il "velo dell'imperscrutabilità tecnica". Saremo in grado di capire ciò che sta accadendo?

Parole chiave: Algoritmi stocastici, Macchine consapevoli, Impero tecnologico, Spiegabilità, Approccio europeo.

The learning deep generative models impose a new vision of AI in relation to autonomy, performance and self-transformation capability. The enormous data assimilation and the continuous capacity for re-processing complexified the complexity. Each device develops others and then still others until a structural concatenation that tends to become a world of its own. So it is increasingly difficult to break the "veil of technical inscrutability". Will we be able to conceive of what is happening?

Keywords: Stochastic algorithms, Conscious machines, Tech empire, Explainability, European approach.

«...lo strumento di un oscuro giocoliere,
di un piccolo torvo pagliaccio che crea mondi nel mondo,
vite nella vita, cervelli nel cervello».

JOHN WILLIAMS, *Nothing But the Night*, 1948

1. L'intelligenza artificiale generativa

Fino a poco tempo fa, per quanto l'uso corrente privilegiasse l'espressione Intelligenza artificiale (AI), gli studi più avveduti preferivano usare il termine al plurale (AIs) per sottolineare la varietà e diversità delle applicazioni, dei modelli di apprendimento (*bottom up* o *top down*, supervisionato, semi-supervisionato, non-supervisionato, per rinforzo), degli algoritmi impiegati, dei problemi sollevati. Si trattava di un orizzonte vasto ed eterogeneo che raccoglieva quelle che venivano definite le tecnologie ingegneristiche e riproduttive, tendenti a imitare e perfezionare, attraverso gli impulsi dei programmatori, le condotte umane, assistendole o sostituendole in un gran numero di contesti. La prospettiva di un'intelligenza artificiale generale (cognitivistica e produttiva), e quindi assimilabile all'intelligenza umana, restava sullo sfondo come l'ipotesi estrema della singolarità tecnologica, preconizzata da Raymond Kurzweil, o della superintelligenza descritta da Nick Bostrom.

Oggi, anche se queste ipotesi sono ancora lontane, i progressi nel *machine learning* hanno dato vita a un nuovo sviluppo, la "Intelligenza artificiale generativa", il mondo di Chat GPT (*General Purpose Technology*), che ha invertito le prospettive di analisi. Il modello ingegneristico e riproduttivo appare marginale rispetto agli sviluppi del mercato e agli indirizzi della ricerca. L'alleanza commerciale tra *big tech* e *startup* consente alle imponenti infrastrutture *hardware* (tra *cloud* e potenza di calcolo) delle prime di accedere alle enormi quantità di dati delle seconde, ampliando indefinitamente le potenzialità delle tecniche di *machine learning* nell'impiegare algoritmi che apprendono direttamente dai dati e sono in grado di eseguire autonomamente un'estrema varietà di compiti, ben al di là degli schemi di programmazione. Una stringa lunga poche centinaia di righe genera, così, automaticamente milioni

di stringhe di codice che si sviluppano per mezzo di reti neurali artificiali riproducendo l'effetto *loop* dei meccanismi ricorsivi del nostro cervello (*deep learning*). «Le applicazioni dell'intelligenza artificiale generativa non sono sistemi di intelligenza artificiale di per sé, ma piuttosto componenti di un sistema generativo più ampio che comprende anche i LLM (*Large Language Model*) utilizzati come materiale di addestramento e i processi che generano i LLM»¹. L'esito di questa procedura, che noi percepiamo come "intelligente", è un *raw executable* che ingloba le migliaia di input umani che hanno fornito quei dati che, filtrati, codificati e assimilati, conducono all'output (ovvero alla decisione o alla valutazione). Si tratta, in prevalenza, di algoritmi stocastici o indeterministici in cui si conosce l'input (ovvero la base dei dati inseriti) ma non è possibile conoscere preventivamente output e dinamica interna.

Le reti neurali artificiali sempre più articolate, che il *business model* (Alphabet, Amazon, Apple, Meta, Microsoft, ecc.) del nostro tempo riesce a simulare, sono in grado di gestire compiti complessi e adattarsi a contesti non strutturati e variabili, come il riconoscimento di immagini o la pluralità delle espressioni linguistiche: «sono gigantesche e nel giro di pochi anni supereranno quelle del cervello umano per numero di neuroni»². Kurzweil, nel proclamare che la singolarità, il superamento meccanico dell'intelligenza umana, è "più vicina", sottolinea che Frontier, il maggior supercomputer al mondo dell'*Oak Ridge National Laboratory*, «può seguire nell'ordine delle 10^{18} operazioni al secondo, che è 10.000 volte superiore della probabile velocità massima di computazione del cervello (10^{14} operazioni al secondo)»³. Già i precedenti supercomputer avevano una capacità di calcolo neppure immaginabile da una mente umana: 200 milioni di miliardi di calcoli al secondo. Quanto potrebbero

¹ * Acknowledge financial support from: PNRR MUR project PE0000013-FAIR. J.E. TARDY, *The Creation of a Conscious Machine. The Quest for Artificial Intelligence*, Mercury Learning and Information, Boston 2024², 180: «The generative AI apps are not AI systems by themselves but rather components of a larger GenAI system that also includes the LLMs used as training material as well as the processes that generate the LLM» (trad. nostra).

² M. SPITZER, *Intelligenza artificiale*, Corbaccio, Milano 2024, 10.

³ R. KURZWEIL, *La singolarità è più vicina*, Apogeo, Milano 2024, 50.

fare tutte le persone sulla Terra se eseguissero un calcolo ogni istante di ogni giorno per 305 giorni. Per non dire che Sycamore, il computer quantistico annunciato da Google, impiega 53 qubit per un totale di 253 valori pari a più di 10 miliardi di combinazioni. Dovrebbe in 200 secondi effettuare quelle operazioni che gli attuali supercomputer all'avanguardia possono compiere in 10.000 anni.

«Ecco la tecnologia definitiva incontrollata e incontrollabile. La verità è che nessuno sa quando, se o esattamente come le IA potrebbero sfuggirci e cosa succederà dopo, nessuno sa quando o se diventeranno completamente autonome o come farle comportare con consapevolezza e allineamento ai nostri valori [...]»⁴.

Questa enorme potenzialità tecnologica consente, per la prima volta nella storia dell'umanità, non solo l'esistenza di macchine che si sviluppano ed evolvono con livelli di autonomia sempre più ampi e spesso assoluti, ma che entrano a contatto, attraverso il *cloud*, con altre macchine, scambiandosi informazioni, creando connessioni operative, perfezionandosi reciprocamente. Si determina quella che già nel 1964 Günther Anders aveva definito l'effetto "imperialistico" o "espansionistico"⁵ delle tecnologie, che tendono a creare un proprio regno "coloniale" di servizi e funzioni.

Accanto a questo effetto espansionistico dobbiamo registrare un effetto "confluenza", per cui una tecnologia rifluisce in un'altra tecnologia, trasferendo gli apprendimenti e producendo effetti ulteriori (*expansive technologies*): «uno strumento in grado di usare altri strumenti»⁶. Ad esempio, AlphaFold2, ideato per giocare a Go, è stato utilizzato per prevedere la struttura delle proteine, un elemento cruciale per la

⁴ M. SULEYMAN-M. BHASKAR, *The Coming Wave*, Penguin Random House, Boston 2023, cap. 7, 21, 22 (versione digitale): «Here is the definitive uncontained and uncountainable technology. The blunt trust is that nobody knows when, if, or exactly how AIs might slip beyond us and what happen next, nobody knows when or if they will become fully autonomous or how to make them behave with awareness of and alignment with our values [...]» (trad. nostra).

⁵ G. ANDERS, *Noi figli di Eichmann*, Giuntina, Firenze 1995, 53-55.

⁶ J. KAPLAN, *Generative A I.*, Luiss, Roma 2024, 191.

progettazione di nuovi farmaci e vaccini. AlphaFold2 esegue in minuti un compito che un tempo richiedeva anni⁷. Gli stessi principi operativi sono stati adottati dalle forze armate statunitensi per pilotare un veicolo da ricognizione U-2. I *Large language model* sono stati collegati alla Risonanza magnetica funzionale per perfezionare le tecniche di scannerizzazione del cervello, ma vengono anche impiegati per un nuovo orizzonte della scienza: gli “animalgoritmi” dell’ecoacustica, gli studi per la comprensione delle comunicazioni animali.

2. Il problema “scatola nera”

Sono solo alcuni dei tanti esempi di una crescita esponenziale delle applicazioni. Secondo Ethan Mollick, arriverà un momento in cui ci renderemo conto che i *Large Language Models* (la nuova forma di intelligenza artificiale generativa) non agiscono come tutti si aspettano che agisca un computer: «invece si comportano piuttosto come una persona»⁸. In effetti non solo usiamo un linguaggio sempre più antropomorfo per cui le macchine imparano, dialogano, capiscono, ma «più le persone attribuiscono loro una soggettività fittizia, più possono essere sfruttate»⁹. È proprio in virtù dell’assimilazione al nostro modo di pensare e alle nostre esigenze che ci attendiamo che l’intelligenza artificiale migliori ogni aspetto della nostra esistenza. Pretendiamo che ci dia i dati corretti di una ricerca, che scriva per noi una lettera, che faccia una diagnosi, che esegua un lavoro, che guidi un’autovettura, che individui un volto, che colpisca un nemico. Insomma vogliamo che faccia quello che faremmo noi, ma meglio. Quanto meglio? Fino a che punto i nostri

⁷ AlphaFold utilizza reti neurali per riconoscere schemi in vasti database di strutture proteiche e sequenze di aminoacidi. Ha predetto la struttura di oltre 200 milioni di proteine, rendendo i dati liberamente accessibili attraverso il database di strutture proteiche AlphaFold.

⁸ E. MOLLICK, *Co-intelligence. Living and Working with AI*, Penguin Random House, London, 2024, xi: «instead they act more like» (trad. nostra).

⁹ *Ibid.*, 56: «the more false agency people ascribe to them, the more they can be exploited»..

parametri di giudizio possono o devono rimanere allineati con i suoi? Il continuo crescere della complessità degli algoritmi e delle funzioni che affidiamo loro non potrebbe determinare delle intelligenze "altre", degli artefatti cognitivi, delle menti esotiche, degli agenti generativi?

Il problema non è quanto l'intelligenza artificiale sia umana, ma quanto possiamo inserirla *human in the loop*, quanto possiamo preservare la dimensione umana di una tecnologia che si sviluppa in termini sempre più invasivi, indipendentemente dalla nostra programmazione e dalle nostre capacità di controllo. È il problema dei limiti dell'interpretabilità meccanicistica, il *black box effect*, emerso per caso all'interno della sfida informatica all'intelligenza umana nel gioco, il Go, con il maggior numero di varianti che si conosca. Fa parte ormai della storia il "colpo 37" con cui il computer *AlphaGo* (il progenitore di *AlphaFold2* di cui ho parlato in precedenza) ha sconfitto il campione del mondo, Lee Sedol: «Era diversa da qualunque altra cosa un computer avesse mai fatto prima. Ed era anche diversa da qualunque cosa un essere umano avesse mai preso in considerazione. Era qualcosa di nuovo, una totale rottura con la tradizione, una deviazione radicale da migliaia di anni di esperienza accumulata»¹⁰.

A partire da quel momento ci siamo resi conto che una cappa sempre più spessa di opacità stava penetrando nei nostri sistemi di programmazione e non sapevano come evitarla. Più l'intelligenza artificiale generativa rendeva questi sistemi complessi per rispondere alla diversità delle nostre esigenze, più diminuiva la trasparenza. Il vantaggio tecnologico straordinario e inatteso di giungere là dove il cervello umano non era mai arrivato, di concepire vie nuove per sviluppare la ricerca, ha reso sempre più difficile garantire l'allineamento ai nostri valori e la spiegabilità delle procedure. «Quando la complessità dell'algoritmo da addestrare è cresciuta, è cresciuta di pari passo l'opacità [...]. I sistemi generativi sono inspiegabili, disallineati e con i piedi per aria»¹¹.

¹⁰ B. LABATUT, *Maniac*, Adelphi, Milano 2023, 314.

¹¹ T. POGGIO-M. MAGRINI, *Cervelli menti algoritmi. Il mistero dell'intelligenza naturale, gli enigmi di quella artificiale*, Sperling & Kupfer, Milano 2022 (ed. digitale), 167.

Abbiamo un assommarsi di problemi tecnici, sistemici, giuridici, filosofici, a cui si aggiungono le scelte politiche di fondo di quello che è stato definito il *Tech Coup*, il colpo di stato tecnologico avvenuto nella Silicon Valley attraverso il connubio tra vertici militari, finanza e imprese informatiche¹². Negli algoritmi indeterministici o stocastici, propri dei sistemi di intelligenza artificiale generativa, la computazione è l'esito dei parametri e delle scelte casuali dell'algoritmo stesso per cui gli stessi programmatori non sanno a che risultato perverranno e perché:

«I modelli di apprendimento automatico basati sui dati [...] sono scatole nere; possiamo interagire con loro ma non sbirciare al loro interno. Tuttavia, non è solo l'IA all'avanguardia a essere incomprensibile. Si dice che nemmeno gli ingegneri di Google siano in grado di comprendere appieno il funzionamento interno del loro motore di ricerca, sul quale molti di noi fanno affidamento per prendere le proprie decisioni quotidiane»¹³.

Gli algoritmi indeterministici stanno costruendo una rete di correlazioni tanto inesplicabili quanto efficienti, aprendo l'orizzonte a nuovi percorsi logici e forse a una vera e propria rivoluzione cognitiva. Se le reti neurali artificiali imitano il nostro cervello, non possono evitare quel *Bereitschaftspotential* di cui parlano i neurologi e di cui ci rendiamo conto tutte le volte in cui avvertiamo di non essere padroni in casa nostra, come insegnava Freud.

Perché le reti neurali artificiali non dovrebbero essere opache come sono opachi i nostri pensieri?

¹² M. SCHAAKE, *The Tech Coup. How to save Democracy from Silicon Valley*, Princeton University Press, Princeton/Oxford 2024.

¹³ U. GASSER-V. MAYER-SCHÖNBERGER, *Guardrails. Guiding Human Decisions in the Age of AI*, Princeton University Press, Princeton & Oxford, 2024, 13 «Data-driven machine learning models... are black boxes; we can interact with them but not peek into them easily. It's not just cutting-edge AI that is incomprehensible, however. Infamously, it is said that even Google's own engineers can fully understand the inner workings of their search engine, upon which so many of us rely to inform our daily decisions» (trad. nostra).

Al di là del diffuso senso di apprensione, «sempre più persone provano il medesimo disagio»¹⁴, sollevato da questo inconscio digitale, ed emerge il sospetto che ci stiamo avviando verso una sistematica incoscienza digitale. Tecnicamente abbiamo imboccato una via apparentemente senza uscita, perché allo stato delle nostre conoscenze non siamo in grado di aprire la scatola nera per guardarci dentro. Le varie tecniche di risoluzione di questi problemi che sono in corso di sperimentazione (*clustering, forest forecasting, self-explainable artificial agents, built on top or built within, off switches, eliciting latent knowledge, testing with concept activation vectors*) non pare forniscano risultati soddisfacenti per tutte le applicazioni e per tutte le funzioni.

A questa opacità tecnica dovremmo aggiungere quella che potremmo chiamare l'opacità sistemica. Non possiamo controllare gli "effetti di rete": aperta una scatola nera se ne trova un'altra e poi ancora un'altra. La complessità delle funzioni dell'intelligenza artificiale si complessifica. Spesso non siamo neppure in grado di capire quando gli algoritmi deterministici mutino in un algoritmo stocastico, come è avvenuto nel caso del "colpo 37". L'effetto domino del paradigma della confluenza e l'effetto ibridazione del paradigma della convergenza creano un labirinto matematico da cui è difficile venir fuori. «La computazione si nasconde sempre più nel tessuto stesso di ogni oggetto che ci circonda; e, vista la sua inarrestabile espansione, opacità e imprevedibilità non fanno che aumentare»¹⁵.

Non dovremmo meravigliarci di tutto questo perché la struttura matematica dell'intelligenza artificiale ci pone di fronte a quanto ci hanno insegnato Gödel e Neumann: ogni sistema matematico è strutturalmente incompleto e tendenzialmente incoerente per cui «non esiste un algoritmo per risolvere il problema della fermata, per decidere se un dato

¹⁴ B. CHRISTIAN, *The Alignment Problem. How Can Artificial Intelligence Learn Human Values*, Atlantic Books, London 2021, 87: «a growing number are feeling that same unease» (trad. nostra).

¹⁵ J. BRIDLE, *Nuova Era Oscura, Nero*, Roma 2023, 215 (edizione digitale).

programma si fermerà mai»¹⁶. La vertigine estrema di un calcolo infinito. La vertigine che Gödel affronta in una scarna (appena una paginetta) equazione che poi è stata pubblicata con il titolo *La prova matematica dell'esistenza di Dio*. Solo la matematica ci consente di intuire l'infinito e di esprimerlo graficamente? L'immagine galileiana di un mondo fatto di numeri ed equazioni si ripresenta nella forma «tutto è software, Dio è un programmatore di computer... Tutto è algoritmo»¹⁷. Un'immagine che potrebbe tanto giustificare letture materialistiche quanto suggerire tendenze mistiche come quelle che Federico Faggin espone da tempo nei suoi libri¹⁸ oppure come l'idea di un "principio cosmogonico", di un'intelligenza artificiale/spirito santo delineata nell'"ecoteologia" di Leonardo Boff¹⁹.

3. La politica dell'opacità

Addentrarci nel mistero dell'esistenza non è il modo migliore per aprire la scatola nera. Semmai ci rimanda al problema dell'approccio filosofico: come entrare a contatto con queste entità intelligenti, con queste menti esotiche, che abbiamo creato e che ormai ci sfuggono? Dobbiamo continuare a seguire la via antropomorfa che riproduce i modelli della nostra intelligenza, abituandoci all'idea di dovere, un domani, dialogare con la "loro" coscienza e accettare la "loro" moralità? Macchine di Dio? Macchine sapienti? Macchine coscienti? Macchine ingannevoli? Dobbiamo addentrarci nel "loro" mondo matematico, cercando gli algoritmi "giusti"? Gert Gigerenzer, ad es., sostiene che non è affatto vero che l'unica via possibile per tener dietro alla complessità

¹⁶ G. J. CHAITIN, *Alla ricerca di Omega*, Adelphi, Milano 2007, 155.

¹⁷ G.J. CHAITIN, *Teoria algoritmica della complessità*, Giappichelli, Torino, 2006, 36.

¹⁸ Cf. da ultimo F. FAGGIN, *Oltre l'invisibile. Dove scienza e spiritualità si uniscono*, Mondadori, Milano 2024.

¹⁹ L. BOFF, *Soffia dove vuole. Lo Spirito Santo dal Big Bang alla liberazione degli oppressi*, EMI, Verona 2019.

degli sviluppi, garantendo l'accuratezza degli esiti, sia necessariamente segnata dall'opacità. A suo avviso, «in condizioni di incertezza, gli algoritmi trasparenti sono spesso altrettanto accurati degli algoritmi della scatola nera»²⁰.

Si tratta allora di una scelta politica? Quella scelta politica che troviamo all'origine del web, quando si è deciso di non tenere traccia della provenienza dei bit, dell'identità dei programmatori e delle loro intenzioni? Il web è stato fatto per ricordare tutto, tranne la sua origine, per cui si dà per scontato che sia strutturalmente decontestualizzato e quindi tendenzialmente indecifrabile. Una nuova forma di *arcana imperii* che, nascosta e strutturalmente impalpabile, si appropria progressivamente del controllo della società. L'intelligenza artificiale aggiunge all'opacità tecnica del *black box model* l'imperscrutabilità della gestione dei dati e delle informazioni, della sede dei *server* o degli *host*, e degli stessi meccanismi della *governance* di tutta questa nuova dimensione del potere. Esiste uno stretto rapporto tra l'infosfera in cui ci muoviamo, l'infodemia che ci condiziona, e l'infocrazia che ci governa. Più la prima diventa indispensabile per il funzionamento della nostra società, meno sappiamo come i dati vengono prodotti, elaborati e impiegati.

Del resto sono opachi tanti altri aspetti della società in cui viviamo. Frank Pasquale descrive la nostra *black box society*²¹, sottolineando tre forme di opacità di cui la nostra organizzazione sociale non pare possa fare a meno. *Real secrecy*: quando si limita l'accesso a determinate informazioni o a determinate persone (Pin, password). *Legal secrecy*: l'obbligo di mantenere segrete determinate informazioni (le normative sulla privacy, il segreto militare, il segreto commerciale). *Obfuscation*: il comportamento deliberato per rendere difficile l'acquisizione di determinati dati (ad es. quando un'impresa trasmette decine di pagine di documenti, rendendo la ricerca degli elementi essenziali un ago nel pagliaio).

²⁰ G. GIGERENZER, *Perché l'intelligenza umana batte ancora gli algoritmi*, Raffaello Cortina, Milano 2023, 188.

²¹ F. PASQUALE, *The Black Box Society. The Secret Algorithms that Control Money and Information*, Harvard University Press, Harvard 2016, 6.

Abbiamo un continuo sovrapporsi dell'opacità intenzionale, che elaboriamo consapevolmente per la tutela di diritti e interessi o per eludere vincoli e controlli, alla opacità tecnologica, che in parte subiamo, per l'incapacità di spiegare gli sviluppi matematici ad alta dimensionalità e il loro impiego nei sistemi di *machine learning*, e in parte sfruttiamo proprio per queste caratteristiche. A tutto questo va unito il fatto che *startup* e *big tech* si sviluppano all'interno del contesto economico e del *trading* finanziario che sono strutturalmente opachi, perché sono il risultato di una pluralità estremamente eterogenea di impulsi, ordini, reazioni, interazioni, aspettative...

Gli interessi del mercato e della politica hanno imposto di erigere il segreto industriale e il segreto militare al di sopra della tutela dei diritti fondamentali. OpenAI si rifiuta di divulgare i dettagli di come ha progettato e addestrato GPT-4, accampando la necessità di difendersi dalla concorrenza di Google e delle altre aziende, in particolare cinesi. PayPal non è tenuta a rivelare i propri algoritmi di concessione del credito. COMPAS ha negato l'accesso alle regole di funzionamento degli algoritmi sulla valutazione della pericolosità sociale e i giudici gli hanno dato ragione²². Altrettanto occulti sono i software bancari che gestiscono la negoziazione algoritmica ad alta frequenza²³. Il segreto industriale è un elemento essenziale del nostro modello di sviluppo e non può essere smontato facilmente. Sempre più spesso ci troviamo di fronte a una scelta senza alternative: o si proteggono gli interessi economici o si tutelano i diritti fondamentali. In genere, la scelta è a favore dei primi anche perché «capitalismo e algoritmi sono consustanziali», l'algoritmo «è interamente catturato [...] nella quantificazione capitalistica»²⁴.

Ancora più rigido è il segreto sul ruolo dell'intelligenza artificiale nella produzione di armi sempre più letali, nella ricerca genetica e batteriologica, nella gestione delle strategie nucleari, negli attacchi ai sistemi

²² *State v. Loomis* 881 N.W.2d 749 (Wis. 2016)

²³ *Singapore International Commercial Court of The Republic of Singapore* [2019] Sghc(I) 03.

²⁴ M. BENASAYAG-A. PENNISI, *ChatGPT non pensa (e il cervello neppure)*, Jaca Book, Milano 2024, 108.

informatici. Non solo la ragion di Stato prevale incondizionatamente sulla trasparenza, ma «una parte importante dell'utilità delle armi cibernetiche deriva proprio dalla loro opacità: il loro disvelamento può effettivamente ridurre parte della forza»²⁵. Il problema è aggravato dal fatto che non solo manca una qualsiasi ipotesi di governo mondiale di questi fenomeni (qualcosa di simile alla limitazione della proliferazione nucleare), ma non è neppure ipotizzabile se pensiamo al moltiplicarsi dei conflitti e all'acuirsi delle tensioni tra Nord e Sud del mondo o meglio tra sfera di influenza americana e Brics.

Va inoltre tenuto presente che il panorama geopolitico è cambiato²⁶. Sono emerse nuove entità, le big tech, che condizionano la politica, alterano i rapporti economici, intervengono nelle guerre, controllano il flusso dei dati e li rielaborano in base alle proprie esigenze. Sono quanto di più oscuro vi sia nel mondo oscuro dell'intelligenza artificiale. *Tech States? Cloud Empires?* Non sappiamo chi governa effettivamente questi nuovi attori privati della proliferazione tecnologica e quindi ogni pretesa di introdurre vincoli di trasparenza corre il rischio di sorreggere «a tempo indeterminato, istituzioni vuote, con il risultato di una dipendenza anziché di una democrazia»²⁷.

4. I timidi passi del diritto

Come avviene sempre nella storia dell'umanità, c'è un vuoto tra l'effettività politica e la teoria politica. Mentre la prima si arrocca dentro le nuove forme di *arcana imperii*, l'esigenza di trasparenza ricorre sistematicamente nelle dichiarazioni internazionali e nelle stesse linee guida delle maggiori imprese. Ad es., un recente documento dell'IBM

²⁵ H.A. KISSINGER-E. SCHMIDT-D. HUTTENLOCHER, *L'era dell'Intelligenza artificiale. Il futuro dell'identità umana*, Mondadori, Milano 2023, 131.

²⁶ A. MHALLA, *Technopolitique. Comment la technologie fait de nous des soldats*, Seuil, Paris 2024.

²⁷ V. LEHDONVIRTA, *Cloud Empires. Come le piattaforme digitali stanno superando gli Stati e come possiamo riprendere il controllo*, Einaudi, Torino 2023, 268.

dichiara «Imperceptible AI is not ethical AI», mentre l'Alto Commissario ONU per i Diritti Umani, nel Rapporto del 13.09.2021 su *Il diritto alla privacy nell'era digitale*, ha rimarcato con forza che l'opacità rende difficile valutare in modo significativo gli effetti dei sistemi di intelligenza artificiale sulla tutela dei diritti fondamentali. Nel 2020 il *White Paper* della Commissione europea *On Artificial Intelligence. A European approach to excellence and trust* affermava che «le caratteristiche specifiche di molte tecnologie di IA, tra cui l'opacità (*black box effect*), la complessità, l'imprevedibilità e il comportamento parzialmente autonomo, possono rendere difficile la verifica del rispetto delle norme del diritto dell'Unione europea in vigore, e ostacolarne l'effettiva applicazione, al fine di proteggere i diritti fondamentali».

La trasparenza, del resto, è uno dei temi fondamentali dei due interventi normativi più importanti sull'intelligenza artificiale dell'Unione europea: il GDPR (Regolamento UE 2016/679) e l'*AI Act* (Regolamento UE 2024/1689). Documenti che costituiscono il primo tentativo, da parte di un grande organismo politico, di mettere ordine al quadro caotico degli sviluppi tecnologici. Va detto che il GDPR procede con molta cautela sul diritto a ottenere una spiegazione della decisione conseguita, sottolineando che non deve «incidere negativamente» sul segreto commerciale. Nell'*AI Act* la disciplina è più dettagliata con l'apprezzabile distinzione tra *user-empowering explainability* e *compliance-oriented explainability*, e con il tentativo di instaurare un sistema di *guardrail*, stabilendo una gradazione nei livelli di rischio.

La spinta europea sta producendo qualche effetto, se nel 2024 diversi paesi, tra cui gli Stati Uniti, hanno sottoscritto la convenzione quadro del Consiglio d'Europa su *Intelligenza artificiale, Diritti umani, Democrazia e Stato di diritto*. È il primo accordo internazionale giuridicamente vincolante su questi temi e, significativamente, pone all'art. 8 l'esigenza della trasparenza «in relazione alle attività del ciclo di vita dei sistemi di intelligenza artificiale».

Il 30 ottobre del 2023 il presidente degli Usa Joe Biden aveva emesso un *Executive Order* su *The Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, in cui si affermava che «è inoltre

necessario affrontare i rischi più pressanti per la sicurezza dei sistemi di IA – anche per quanto riguarda le biotecnologie, la sicurezza informatica, le infrastrutture critiche e altri pericoli per la sicurezza nazionale – pur tenendo conto (*while navigating*) dell'opacità e della complessità dell'IA». Come interpretare il *while navigating*? Nel senso del rassegnato realismo del documento (*Artificial Intelligence, AI in the UK: ready, willing and able?*) dell'*House of Lords Select Committee* del 2023? Un documento che, sul problema della trasparenza, non esita ad affermare che «sulla base delle prove che abbiamo ricevuto, riteniamo che il raggiungimento di una piena trasparenza tecnica sia difficile, e forse addirittura impossibile, per alcuni tipi di sistemi di IA in uso oggi, e in ogni caso non sarebbe appropriata o utile in molti casi».

Penso che dovremmo prestare particolare attenzione a queste caute inquietudini, perché ci lasciano intuire la sottile crepa che potrebbe fare franare l'effettiva applicazione del principio di esplicabilità. Se inteso nella sua pienezza, questo principio dovrebbe fare in modo che vi siano «esperti e operatori di IA che abbiano il controllo, sappiano cosa stanno facendo e siano in grado e disposti a comunicare, spiegare e motivare ciò che stanno facendo a soggetti umani e non umani»²⁸. Ogni sistema operativo dovrebbe rispecchiare la sintesi «Five W's and How»: vale a dire consentire di individuare *What, Who, When, Why, Where, and How*. Però... *Who* e *What*, "Chi" e "Cosa", sono inestricabilmente confusi nella nebulosa algoritmica perché «the purpose of a system is what it does» (è solo l'esito quello che conta). *When* e *Where*, "Quando" e "Dove", non sono decifrabili nella globalizzazione algoritmica, in cui tutto è filtrato e smaterializzato. *Why* e *How*, "Perché" e "Come", sono interrogativi vuoti nel connessionismo algoritmico costituito dall'impenetrabile intersecarsi di diramazioni e aggregazioni. E allora? Dovremmo aggiungere una sesta W: *While*, ossia *While*

²⁸ M. COECKELBERGH, *Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability in Science and Engineering Ethics* 26 (2020) 268: «society deserves AI experts and operators who are in control, know what they are doing, and are able and willing to communicate, explain and give reasons for what they are doing to human and nonhuman moral patients» (trad. nostra).

navigating, come suggerisce l'*Executive Order* di Biden. Siamo nel mezzo di un percorso e non sappiamo dove ci porterà: «nessuno sa quando, e nemmeno se queste scatole nere assumeranno un colore grigio più gestibile [...] ed è probabilmente un sogno che non si realizzerà mai»²⁹.

²⁹ M.A. BODEN, *Intelligenza artificiale*, in J. AL-KHALILI, *Il futuro che verrà. Quello che gli scienziati possono prevedere*, Bollati Boringhieri, Torino 2018, 127.

LEGALITÀ ARTIFICIALE

Alberto M. Gambino

ABSTRACT

L'adozione dell'IA, senza regolamentazione giuridica adeguata e monitoraggio costante, può mettere a rischio la certezza del diritto e la protezione della persona, delegando la determinazione delle norme a entità non politiche. È urgente un intervento più ampio e cogente del Regolamento europeo sull'IA, affinché la tecnologia non comprometta i principi di giustizia e legalità che sono alla base della nostra società democratica.

Parole chiave: legalità artificiale, sovranità, lex mercatoria, consumerismo.

The adoption of AI, without adequate legal regulation and constant monitoring, can jeopardize legal certainty and the protection of the person, delegating the determination of rules to non-political entities. A broader and more binding intervention of the European Regulation on AI is urgently needed, so that technology does not compromise the principles of justice and legality that are the basis of our democratic society.

Keywords: artificial legality, sovereignty, lex mercatoria, consumerism.

Il tema della “*legalità artificiale*” si propone come una chiave di lettura delle attuali dinamiche giuridiche, in cui le decisioni dispositive (quelle dotate di forza vincolante nei confronti dei destinatari) vengono progressivamente determinate da software, algoritmi e sistemi di intelligenza artificiale (IA)¹. Il sintagma assume un significato centrale in un contesto normativo sempre più interconnesso con soluzioni tecnologiche, che sfuggono in parte al controllo diretto della legislazione formale,

¹ Trascrizione dalla videoregistrazione dell'intervento nel convegno, rivista e autorizzata dall'Autore.

per essere invece modellate da prassi, protocolli e codifiche algoritmiche sviluppate da entità private o pubbliche, ma non dal legislatore.

L'assenza di un testo normativo esplicito, a favore di soluzioni regolate da procedure algoritmiche o software, è un fenomeno che riguarda principalmente il comportamento economico e che si espande rapidamente in altre aree, come quella giuridica e medica. Il fenomeno acquista ulteriore rilevanza alla luce del recente Disegno di legge italiano, che intende regolamentare l'impiego dell'intelligenza artificiale nell'ambito della giurisdizione, ove tale tecnologia potrebbe essere utilizzata per supportare le attività dei magistrati, per esempio, nel contesto della ricerca giurisprudenziale. Gli algoritmi, in effetti, promettono di decodificare non solo il testo delle decisioni, ma anche la logica e il ragionamento sottesi, migliorando così la qualità e l'efficacia delle ricerche giuridiche.

Le applicazioni della IA potrebbero certamente rappresentare un utile ausilio per l'efficienza della giustizia, contribuendo a velocizzare e rendere più precise le operazioni di ricerca. Tuttavia, una questione cruciale sorge quando l'automatizzazione inizia a compromettere il margine di intervento critico del giudice. Un rischio significativo è che l'impiego di algoritmi possa portare a una *standardizzazione delle decisioni* giuridiche, con il rischio che i giudici non esercitino più il necessario vaglio critico delle soluzioni proposte dal software, affidandosi passivamente ai risultati ottenuti.

Tale tendenza potrebbe accentuarsi ulteriormente nei casi di decisioni ripetitive, dove il software, formulando una sentenza "tipo", potrebbe diventare il punto di riferimento decisionale senza un'adeguata riflessione critica da parte del magistrato. Questa dinamica comporterebbe la trasformazione di un processo decisionale complesso in una mera applicazione di regole prestabilite da un algoritmo, con il rischio di compromettere l'autonomia e l'indipendenza del giudice.

Un parallelo interessante si riscontra nel settore *universitario*, quando i professori si confrontano con piattaforme digitali (come il Cineca) per l'accREDITAMENTO dei corsi. In tali contesti, le norme ministeriali sono codificate da programmatori in algoritmi, ma queste, per loro natura, sono

parziali e non rispecchiano pienamente il quadro normativo. Pertanto, quando il software non restituirà un esito positivo, i docenti potrebbero intraprendere un processo di “tentativi” per conformarsi alle aspettative del sistema. Tuttavia, questa prassi solleva interrogativi importanti: *l’algoritmo ha correttamente interpretato la norma*, o è stato travisato in fase di programmazione? La giurisprudenza amministrativa ha stabilito che, in presenza di errori imputabili all’algoritmo, i titoli accademici conseguiti devono essere riconosciuti validi, considerando l’affidamento legittimo degli studenti nell’operato dell’istituzione. Tuttavia, questa “*legalità artificiale*” non corrisponde alla legalità tradizionalmente sancita nelle leggi pubblicate in Gazzetta Ufficiale.

Questo fenomeno assume una rilevanza particolare nell’ambito della *medicina*, dove l’adozione di protocolli medici (spesso influenzati da software e algoritmi) comporta il rischio di una medicina difensiva. I medici, infatti, sono incentivati a seguire scrupolosamente le linee guida ufficiali, per evitare responsabilità legali in caso di esito negativo. La questione si fa ancor più critica quando, di fronte a situazioni complesse, il medico si trova a scegliere tra l’applicazione di un protocollo standardizzato, ma ritenuto inefficace, e una pratica innovativa basata sul suo giudizio clinico, rischiando di esporsi a procedimenti legali per negligenza, qualora il risultato fosse sfavorevole.

Tale “*legalità artificiale*” si estende anche ai settori economici, come quello delle piattaforme digitali, dove *le scelte normative* sono spesso dettate da logiche economiche piuttosto che giuridiche. Il Regolamento europeo sull’IA, ad esempio, si fonda più su una base economica che giuridica, mirando a favorire l’innovazione, piuttosto che a garantire la protezione dei consumatori. Sebbene le normative esistenti offrano una protezione nei confronti delle violazioni dei diritti umani (ad esempio, in materia di scoring creditizio), la regolamentazione dell’IA rimane fortemente orientata verso la promozione dello sviluppo, con un metodo “*deterrente*” piuttosto che preventivo.

Questa impostazione, pur trovando fondamento nelle necessità di sviluppo e di globalizzazione, comporta un’intrinseca “*tolleranza giuridica*” che si allontana dalle rigide tutele previste dagli ordinamenti

statali, con potenziali rischi per i diritti fondamentali delle persone, in particolare per quanto concerne la tutela della privacy e la protezione dei dati personali. L'affidamento di *scelte giuridiche* a un software non regolato in modo chiaro e preciso può portare, infatti, a una violazione dei diritti economici e persino dei diritti umani, considerando che il patrimonio e l'autonomia economica sono componenti essenziali della persona umana.

In definitiva, l'ascesa della *legalità artificiale* si configura come un fenomeno che, se non adeguatamente monitorato e regolato, rischia di minare le basi stesse della *sovranità statale* e del modello di giurisdizione, in quanto la *normativa giuridica* tende a subire l'influenza di codici auto-regolamentati da soggetti imprenditoriali, i quali dettano regole che possono essere intrise di *eticità*, ma non necessariamente di legittimità. In questo quadro, l'IA rischia di determinare un sistema giuridico meno garantista, dove le scelte decisionali si basano sempre più sulle logiche di mercato e sull'efficienza economica, a scapito della tutela dei diritti fondamentali.

Pertanto, la *legalità artificiale* solleva interrogativi cruciali: è giusto che le decisioni normative e dispositive vengano determinate da software e algoritmi privi di una chiara legittimazione giuridica? L'adozione indiscriminata dell'IA, senza una regolamentazione giuridica adeguata e un monitoraggio costante, potrebbe mettere in pericolo i principi di *certezza del diritto* e di *protezione dei diritti fondamentali*, riducendo la fiducia nel sistema giuridico tradizionale e delegando a entità non politiche la determinazione delle norme. È quindi urgente un intervento legislativo, più ampio e cogente del Regolamento europeo sulla IA, che garantisca una supervisione adeguata e una *vigilanza umana* sui processi decisionali alimentati dall'IA, affinché la tecnologia non comprometta i principi di *giustizia* e *legalità* che sono alla base della nostra società democratica.

INTELLIGENZA ARTIFICIALE COME SPECCHIO DELL'UMANO. TRA CONSAPEVOLEZZA DI SÉ E DISTORSIONI MORFOLOGICHE

Andrea Galluzzi

ABSTRACT

Le tecnologie riflettono ciò che siamo, come in uno specchio. Questa dinamica appare oggi evidente grazie allo sviluppo di intelligenze artificiali dotate di capacità "creative" che legano ancora più strettamente il dato tecnologico a quello umano. Tuttavia, se da una parte i sistemi IA rispondono al desiderio umano di potenziamento, dall'altra il loro impiego solleva numerosi interrogativi. In questa epoca di cambiamenti è possibile trovare elementi su cui riflettere per comprendere l'incommensurabilità dell'umano al di là delle sue simulazioni?

Parole chiave: IA, Identità, Coscienza, Sacro, Immagine, Consapevolezza di sé.

The technologies reflect who we are, as in a mirror. This dynamic is evident today thanks to the development of artificial intelligences endowed with "creative" capacities that bind the technological datum even more closely to the human datum. However, while on the one hand, AI systems respond to the human desire for empowerment, on the other their often ambivalent use raises numerous questions. In this era of change, is it possible to find elements on which to reflect in order to understand the incommensurability of the human beyond its simulations?

Keywords: AI, Identity, Consciousness, Sacred, Image, Self-awareness.

1. L'uomo allo specchio (e allo schermo)

Nella sua profonda analisi sull'uomo e la tecnica, Arnold Gehlen (1904-1976) affermava che la tecnica è insita nell'essenza stessa dell'uomo: «in tutto quanto ha di ingegnoso, di semplicemente inverosimile, e nel suo rapporto indiretto con la natura, la tecnica è un vero specchio dell'essere

umano»¹. Il tema dell'uomo allo specchio è ricco di significati simbolici e filosofici. Riguarda l'identità, l'autoconsapevolezza, l'introspezione, il confronto tra l'essere e l'apparire. Allo specchio ognuno vede se stesso, o almeno ciò che appare. Allo specchio ci si osserva, si cercano difetti per cercare di correggerli, si cerca una conformità fra l'immagine che vediamo e l'idea che abbiamo di noi stessi. Così è anche, trasponendo il concetto, per le cose che creiamo con le nostre mani e il nostro ingegno. Sembra prevalere un'attitudine innata nel ritrovare un po' di sé nelle cose che realizziamo e che rispecchiano la nostra personalità, le nostre intenzioni, il nostro modo di essere e di concepire l'esistenza. Se è vero che ogni strumento rimanda ad una certa immagine di noi stessi, possiamo dire – con Adriano Fabris – che macchine e automatismi possono essere considerati «come ciò che riflette la nostra immagine e la prolunga nel mondo»².

Ma che cos'è una "immagine"? Heidegger, riflettendo sull'immagine del mondo moderno, afferma che essa «non indica una mera copia, un calco, ma una forma»³. Essere "in-formati" su qualcosa significa conoscere ciò che abbiamo davanti, ma anche avere coscienza di ciò che sta in esso, del suo essere strutturato a sistema. Nell'era tecnologica macchine e algoritmi entrano a comporre gran parte del quadro di questa immagine e il nostro rispecchiarci in questo gigantesco sistema contribuisce a definire non solo ciò che siamo, ma anche ciò a cui desideriamo uniformarci.

In questo scenario, lo sviluppo dell'intelligenza artificiale (IA), caratterizzato negli ultimi anni da una crescente capacità di linguaggio e generazione di contenuti creativi, ha portato questa esperienza a un livello inedito. Attraverso la metafora dello specchio possiamo concepire l'IA come una tecnologia capace di "proiettare" forme umane anche su ciò

¹ A. GEHLEN, *Die Seele in Technischen Zeitalter* (1957), tr. it.: *L'uomo nell'era della tecnica. Problemi socio-psicologici della civiltà industriale*, Armando, Roma 2003, 33.

² A. FABRIS, *La filosofia e lo specchio delle macchine*, in *In Circolo. Rivista di filosofia e culture* 6 (2018) 28-38, 30.

³ M. HEIDEGGER, *L'epoca dell'immagine del mondo*, in *Id.*, *Holzwege. Sentieri erranti nella selva*, Bompiani, Milano 2002, 107.

che umano non è, restituendo all'osservatore un nuovo modo di interagire con la macchina e oggettivarsi in essa. La naturale tensione a rispecchiarsi nelle proprie creazioni o a trovare qualcosa di sé negli strumenti o negli apparati che utilizziamo, si arricchisce così di nuovi elementi. Grazie all'IA lo strumento diventa un interlocutore, un nuovo compagno di lavoro al quale chiedere consiglio, col quale discutere, nel quale trovare sicurezza, riparo, conforto o – al limite – volere assomigliare.

Nell'ambiente digitale, sempre più mediato dalle IA, allo specchio si sostituisce lo schermo, attraverso il quale passano informazioni, immagini e stimoli che plasmano la nostra percezione di realtà. La pervasività di questi fenomeni spinge a cercare categorie di pensiero o neologismi in grado di catturare l'essenza di queste trasformazioni⁴. Così, ad es., è stato coniato il termine *guscio tecnologico*⁵ per definire tutto l'insieme di tecnologie dell'informazione al servizio della persona, che si interpongono fra il soggetto e il mondo circostante⁶. Questo *guscio* dirige la sua azione proattiva non solo verso il suo interno, cioè verso la persona di cui avere cura, ma anche verso l'esterno, fungendo da meccanismo di accoglienza, difesa o filtraggio, dando al soggetto la possibilità di costruire nuove definizioni di sé. In altre parole, il *guscio tecnologico* non solo ci protegge, ma filtra la nostra percezione, mediando la visione del mondo attraverso i suoi dispositivi. Attraverso di esso la realtà viene elaborata, modificata, enfatizzata e riproposta all'osservatore, il quale, a sua volta, può reinventare e proporre nuove immagini di sé. Esso diventa così un potente strumento di auto-espressione e

⁴ Così è stato, ad esempio, per i termini "Onlife" e "Algor-etica", che indicano rispettivamente la fusione in atto tra la vita online e quella reale, e la necessità di un'etica specifica per gli algoritmi che regolano il mondo. Ma se ne potrebbero citare molti altri.

⁵ Cf. A. GALLUZZI, *Intelligenza Artificiale come "tecnologia del sé"*, in G. PIAIA-R. PRETE-L. STEFANUTTI (curr.), *Intelligenza artificiale e persona umana. Implicazioni etico-giuridiche*, Triveneto Theology Press, Padova 2024, 45-62.

⁶ Rientrano all'interno di questa definizione i social network, i dispositivi indossabili in grado di monitorare il nostro stato di salute, gli smartphone e ogni altro dispositivo elettronico in grado di processare informazione con finalità utili alla soddisfazione di qualche tipo di bisogno umano.

auto-rappresentazione, che ci guida verso il compimento di un desiderio innato: il soddisfacimento della volontà di potenza.

2. Verso il potenziamento. IA come “specchio convergente”

Come uno specchio convergente, le tecnologie influiscono sulla natura del nostro agire focalizzandolo non solo sul soddisfacimento di bisogni, ma anche su un desiderio insito da sempre nell'intimo dell'umano: il nostro potenziamento. Fra le esigenze del mondo digitale spicca il bisogno di sintesi, velocità e automazione. Per questo, dopo la rapida diffusione degli LLM (*Large Language Models*)⁷ e delle IA generative, si assisterà in breve tempo alla diffusione su scala globale dei LAM (*Large Action Models*): agenti artificiali in grado di portare a termine autonomamente, al posto nostro, compiti complessi da eseguire via web, interfacciandosi con tutte le altre infrastrutture digitali per procurarci beni o servizi (dagli acquisti alle prenotazioni, dalla ricerca dati alla logistica ecc.)⁸. L'idea alla base di questo sviluppo è quella di delegare flussi di lavoro e compiti automatizzabili ad agenti IA, lasciando agli umani compiti di supervisione o di semplice fruizione dei servizi.

Con lo sviluppo dell'IA, il tema del potenziamento umano attraverso le tecnologie esce dalla sfera materiale per estendersi anche verso quella mentale e psicologica⁹.

⁷ I Large Language Models (LLM) sono sistemi IA progettati per comprendere e generare linguaggio naturale. La loro capacità di generare risposte coerenti e contestualmente rilevanti li rende strumenti utili in diversi ambiti, e spinge la loro implementazione in innumerevoli applicazioni. ChatGPT ne rappresenta un esempio emblematico.

⁸ L'attivazione di sistemi LAM richiede la presenza di un ecosistema altamente automatizzato in grado di implementare la complessità delle richieste, ed è questa la direzione verso la quale si sta muovendo il mondo digitale. A titolo di esempio si veda: T. MASTERMAN et al., *The Landscape of Emerging AI Agent Architectures for Reasoning, Planning, and Tool Calling: A Survey*, 17 aprile 2024, <https://arxiv.org/abs/2404.11584> (accesso 01.10.2024)

⁹ I dispositivi tecnologici orientati al supporto e al potenziamento delle capacità fisiche sono di uso comune. Basti pensare alle macchine dotate di servomeccanismi o alle protesi biomediche come i pace-makers o gli apparecchi acustici. Anche questi artefatti possono

Questa sfera di potenziamento si articola in tre principali tipologie¹⁰: *potenziamento emotivo*, che enfatizza emozioni, stati d'animo, sentimenti¹¹; *potenziamento cognitivo*, che include l'uso di tecnologie per migliorare l'intelligenza, la memoria, le capacità analitiche e la consapevolezza di sé¹²; *potenziamento morale*, che influisce sulle capacità decisionali, sul comportamento, sui giudizi e sull'etica personale¹³. In seguito a questi sviluppi, la natura dell'agire umano subisce inevitabili trasformazioni, e di conseguenza anche l'etica, strettamente legata all'agire, viene messa alla prova da queste forme di potenziamento. A questo riguardo Hans Jonas osserva:

«[...] la mutata natura dell'agire umano richiede anche un mutamento nell'etica: questo non semplicemente nel senso che nuovi oggetti dell'agire hanno ampliato materialmente l'ambito dei casi a cui devono essere applicate le regole vigenti del comportamento, ma nel senso più radicale che la natura qualitativamente nuova di certe nostre azioni ha dischiuso una nuova dimensione eticamente significativa, di cui non esistono precedenti nei criteri e nei canoni dell'etica tradizionale»¹⁴.

Tra le nuove sfide etiche legate allo sviluppo dell'IA spicca il perfezionamento delle tecniche di *allineamento*: termine che indica l'obiettivo di garantire che i sistemi IA forniscano risposte coerenti e adeguate,

essere inseriti concettualmente all'interno dell'insieme del "guscio tecnologico".

¹⁰ Cf. C. TADDEI FERRETTI-A. GINI, *Mente moralmente potenziata secondo le neuroscienze vs sviluppo morale responsabile secondo Lonergan*, in G. GUGLIELMI (cur.), *In dialogo con Bernard J.F. Lonergan. Scritti in onore di Rocco Pititto*, Aracne, Canterano 2018, 225–236: 231.

¹¹ Attraverso strumenti tecnologici, è possibile monitorare e analizzare stati d'animo e sentimenti, favorendo una maggiore stabilità emotiva o la capacità di comprendere e gestire le proprie emozioni e quelle altrui.

¹² Ad es., attraverso appositi algoritmi si possono sviluppare strumenti per aumentare l'efficacia del ragionamento, potenziare la memoria e affinare il pensiero logico-astratto.

¹³ Questo ambito si propone di utilizzare tecnologie per guidare e migliorare le azioni e i comportamenti, stimolando tendenze positive e orientamenti etici.

¹⁴ H. JONAS, *Philosophical essays. From ancient creed to technological man* (1974), tr. it.: *Dalla fede antica all'uomo tecnologico. Saggi filosofici*, il Mulino, Bologna 1991, 41.

prevenendo fenomeni come *allucinazioni* o *bias*¹⁵. Il problema dell'allineamento non è solo di natura scientifica o tecnica, ma anche etica, a causa delle innumerevoli ripercussioni sulla sfera umana. Le domande a cui gli agenti IA sono chiamati a rispondere, infatti, spesso toccano temi complessi legati alla storia, ai valori, alle credenze e alle norme sociali, su cui costruiamo nuove narrazioni. Non si tratta solo di garantire che il sistema raggiunga con efficienza gli obiettivi prefissati, ma anche che le sue risposte riflettano le motivazioni e le aspirazioni umane più profonde, rispettando la diversità culturale e i principi morali universalmente condivisi.

3. Il rischio del disallineamento. IA come “specchio divergente”

Al soddisfacimento della volontà di potenza e ai vantaggi ottenuti grazie alle nuove possibilità tecnologiche si accompagnano spesso una certa difficoltà etica e un senso di disorientamento esistenziale. Questi sentimenti emergono in risposta agli squilibri provati dai rapidi mutamenti indotti dalle tecnologie nella sfera sociale, che talvolta si rilevano profondi e destabilizzanti. Edgar Morin parla ad es. di un *principio di incertezza*¹⁶, frutto «di una dialogica (relazione allo stesso tempo antagonista, concorrente e complementare) fra l'ordine, il disordine e l'organizzazione»¹⁷, ed evidenzia come ogni innovazione possa essere correlata a profonde metamorfosi esistenziali: «Le grandi trasformazioni sono morfogenesi, creatrici di nuove forme, che possono costituire vere metamorfosi. In ogni modo, non esiste evoluzione che non sia

¹⁵ Nell'ambito dell'IA le “allucinazioni” e i *bias* rappresentano due tipi di errori che possono influenzare il comportamento e le risposte di un sistema. Il primo tipo di errore si verifica quando un modello di IA genera informazioni errate o inventate che non trovano riscontro nei dati di addestramento; il secondo è invece costituito da distorsioni sistematiche nelle risposte, derivanti da errori o squilibri presenti già nei dati di addestramento.

¹⁶ Cf. E. MORIN, *Les sept savoirs nécessaires à l'éducation du futur* (2000), tr. it.: *I sette saperi necessari all'educazione del futuro*, Raffaello Cortina Editore, Milano 2001, 86-95.

¹⁷ *Ibid.*, 86.

disorganizzatrice/riorganizzatrice nel suo processo di trasformazione o di metamorfosi»¹⁸.

Queste riflessioni trovano corrispondenza con quelle di altri autori. Ad es. Matteo Ciastellardi, analizzando il rapporto tra lo sviluppo tecnologico e l'interiorità umana, evidenzia come la natura della tecnica si manifesti attraverso un'azione coinvolgente, capace di influenzare il soggetto, scomponendone e ricomponendone l'individualità e la soggettività¹⁹. Egli osserva una sorta di co-implicazione tra gli oggetti d'uso e i loro fruitori, particolarmente intensa quando il legame funzionale è più stretto. Le tecnologie moderne tendono a trasformare questa relazione, passando da una semplice interazione alla "telensione": un termine frutto della crasi tra "tele-controllo" e "tensione", con cui Ciastellardi intende descrivere lo spostamento cognitivo tra l'uomo e i suoi oggetti d'uso, i quali «[innalzano] la soglia di potenza del soggetto, rendendolo al contempo vittima dell'apparato di cui fruisce e a cui si affida»²⁰.

In altre parole le tecnologie moderne, sia quando amplificano enormemente le possibilità espressive dell'individuo, sia quando si configurano come sistemi omologanti e massificanti, tendono a influire profondamente sull'interiorità, modificando l'identità del soggetto e le sue azioni. Il percorso della vita interiore – inteso come cammino esistenziale, maturazione di una visione del mondo o ricerca di senso – sembra così sottoposto a continue trasformazioni, causate dalla *telensione* con apparati di crescente complessità e dai "sovraccarichi" caratteristici dell'era dell'informazione. Si tratta di un effetto particolarmente evidente nell'uso delle IA in vari contesti. Nel corso del XX secolo queste incertezze sembrano essersi trasformate da temporanee a permanenti a causa delle instabilità provocate dai repentini cambiamenti indotti

¹⁸ *Ibid.*, 84.

¹⁹ Cf. M. CIASTELLARDI, *Provocazioni della tecnica*, in P. D'ALESSANDRO-A. POTESTIO (curr.), *Filosofia della tecnica*, LED, Milano 2006, 111-132.

²⁰ *Ibid.*, 118.

dal rapido sviluppo tecnologico²¹. La ricerca di una integrità personale nel continuo polimorfismo a cui ci ha abituato l'orizzonte postmoderno è diventata più complessa. Tale ambiguità sembra spingere verso una sistematica rinuncia a ciò che è stabile e definitivo, in favore di un orizzonte di riferimenti in costante evoluzione.

L'emblema della *società liquida*²², definita con efficacia da Zygmunt Bauman (1925-2017), rappresenta perfettamente gli effetti di queste provvisorietà in rapporto alle forme della vita individuale e sociale. Mutamenti che – come suggerisce lo psicologo Tonino Cantelmi riferendosi al ruolo ambiguo assunto dalla tecnica moderna in questa dinamica esistenziale – portano alla definizione di una condizione *tecno-liquida*²³: un presupposto che «aiuta il tecno-uomo a rimanere aggiornato e connesso»²⁴. L'individuo si trova così diviso tra l'utilizzo di strumenti evoluti (come le IA e i loro agenti) che lo rendono funzionalmente adeguato alla "liquidità" circostante, e il rischio di perdere l'integrità interiore. Tale dissonanza può manifestarsi sotto forma di un *analfabetismo emotivo*²⁵, cioè una difficoltà nel riconoscere e descrivere sentimenti ed emozioni, che tende a incidere profondamente nella struttura della mente umana in senso disorientante. La sensazione di disagio e smarrimento esistenziale che sembra caratterizzare lo sfondo dell'era tecnoliquida potrebbe essere interpretata – suggerisce Cantelmi – come il preludio di un nuovo «passaggio evolutivo», una mutazione antropologica che porterebbe a riorganizzare le capacità neuro-cognitive e a

²¹ Sulla scia del pensiero di Morin sarebbe possibile qui accostare il tema del disorientamento interiore a quello, molto vasto, dell'educazione. Si osserva infatti come il sistema tecnico tenda a privilegiare prestazioni, efficienza, produttività, competenze professionali, trascurando le molteplici caratteristiche che contribuiscono a formare la persona nella sua interezza.

²² Cf. Z. BAUMAN, *Liquid Modernity* (2000), tr. it.: *Modernità liquida*, Laterza, Roma 2012.

²³ Cf. T. CANTELMINI, *Tecnoliquidità. La psicologia ai tempi di Internet: la mente tecnoliquida*, San Paolo, Cinisello Balsamo (Mi) 2013.

²⁴ *Ibid.*, 161.

²⁵ Cf. *Ibid.*, 165.

modificare anche alcuni pilastri nell'identità personale²⁶. Per cercare di definire alcuni elementi di questo passaggio possiamo fare ricorso nuovamente alla metafora dello specchio e pensare a come le tecnologie IA ci spingano sempre più a riflettere sulla nostra stessa identità.

4. Oltre lo specchio. L'IA come "portale" verso il mondo digitale

Nel famoso racconto di Lewis Carrol, *Attraverso lo specchio e quel che Alice vi trovò* (1871), la protagonista attraversa uno specchio che funge da portale tra due mondi, quello reale e quello riflesso, in cui le regole ordinarie del mondo reale vengono sovvertite. Il passaggio di Alice attraverso lo specchio può essere interpretato come una metafora del viaggio interiore e della crescita personale. Lo specchio riflette non solo l'esterno, ma anche l'interno: la coscienza, le paure e i desideri di Alice. Nel suo viaggio Alice si confronta con se stessa e con la sua identità, così come anche noi oggi, di fronte agli sviluppi della tecnologia, siamo chiamati a riflettere su cosa caratterizzi la nostra natura umana. Nel contesto mediato dalle tecnologie, l'IA sembra assumere il ruolo dello specchio di Alice, trasferendo al di là dello specchio/schermo, cioè nel dominio del digitale, concetti che prima erano strettamente legati alla sfera umana o confinati alle narrazioni della fantascienza. L'IA si presenta quindi come un "portale" tra due mondi: quello fisico e quello digitale, in cui si intrecciano e si proiettano reciprocamente speranze, desideri, inclinazioni, sogni e nei quale si sfumano elementi reali e virtuali.

A titolo di esempio possiamo pensare allo sviluppo dei nostri avatar: agenti digitali in grado di rappresentarci, sostituirci e ridefinire la nostra identità e storia²⁷. Non è improbabile immaginare un futuro prossimo in cui queste entità saranno molto diffuse e sopravviveranno a noi stessi

²⁶ Cf. *Ibid.*, 10.

²⁷ A titolo di esempio si veda: J. S. PARK ET ALII, *Generative Agents: Interactive Simulacra of Human Behavior*, 07 aprile 2023, <https://arxiv.org/abs/2304.03442v2> (accesso 01.10.2024).

(come sopravvivono a noi anche i profili social), continuando a simulare la nostra presenza dialogando con le future generazioni sulla base dei dati raccolti in vita e su cui costruire nuove narrazioni. Questo particolare processo di digitalizzazione, o meglio *avatarizzazione* – come suggerisce Andrea Pinotti in una sua ampia e variegata analisi – si presta particolarmente ad essere interpretato come soglia oltre la quale il soggetto può dare nuove definizioni di sé, ma anche maturare di sé nuove comprensioni. Scrive Pinotti:

«Lo spettro funzionale dell'*avatar* va dalla rivelazione del Sé (una sorta di autoritratto veritiero) alla dissimulazione della maschera: in virtù di una gamma così vasta di possibilità, l'*avatar* funge da operatore d'identità che consente un numero virtualmente infinito di negoziazioni. Esso permette quell'esperienza peculiare che è stata denominata "auto-empatia": la possibilità cioè di empatizzare con l'altro-in-me, di assumere la prospettiva sul mondo come da un punto di vista esterno al mio, eppure sempre in qualche modo mio»²⁸.

Per riflettere su alterità ed empatia nel contesto del rapporto con agenti artificiali può essere utile riferirsi alle considerazioni di Theodore Lipps (1851-1914), uno dei padri fondatori della psicologia moderna. Dal suo punto di vista, in una relazione intersoggettiva la *partecipazione emotiva interiore*²⁹ è fondata su un processo cognitivo la cui dinamica non è immediata, ma frutto di un processo di consapevolezza. La consapevolezza sulla realtà è data dall'incontro di tre ambiti conoscitivi che fondano la vita psichica: quello della *percezione sensibile*, della *percezione interna* e dell'*empatia*, coordinati e vagliati secondo le leggi della ragione. Scrive Lipps:

«In questo sistema, secondo le leggi della ragione, le singole affermazioni della percezione sensibile, del ricordo, di quell'auto-oggettivazione, si correggono fra di loro reciprocamente. Ciò che, nelle affermazioni dei tre istinti, non tollera i controlli operati tramite tali leggi, si chiama

²⁸ A. PINOTTI, *Procuratori del sé: dall'avatar all'avatarizzazione*, in T. GATTI-D. MAINI (curr.), *Visual studies. L'avvento di nuovi paradigmi*, Mimesis, Milano 2019, 27-40: 28.

²⁹ Cf. I. ROTELLA, *Altro come sé. Il problema dell'empatia e dell'alterità a partire da Theodor Lipps*, in L. ALICI-F. MIANO (curr.), *L'etica nel futuro*, Orthotes, Napoli 2020, 191-201: 192.

“parvenza” o “illusione”. Ciò che li tollera, proprio passando per tali controlli diventa “verità”. Verità, in tutti gli ambiti della conoscenza della realtà, non è dunque l’affermazione di quel triplice istinto, ma ciascuna delle sue affermazioni che sia stata sanzionata dalla ragione. Così, in particolare, non è che qualsiasi cosa che venga empatizzata e innanzitutto ritenuta reale sia qualcosa di definitivamente o di oggettivamente reale»³⁰.

In altre parole, seguendo Lipps, anche se non si esclude la possibilità di una relazione empatica con soggetti non umani, è possibile cadere in errore. Così come si è soggetti ad illusioni ottiche, è possibile sperimentare *illusioni empatiche*³¹, come nel caso di relazioni con persone apparentemente amiche che però celano intenti malevoli, o come nel caso di interazioni con agenti artificiali, frutto di simulazioni non supportate da vera autocoscienza, e quindi ingannevoli. Secondo Lipps la teoria dell’empatia si basa su una percezione del mondo fondamentalmente soggettiva, tanto da poter affermare – sottolinea Ivan Rotella – che «l’incontro con l’altro finisce dunque per limitarsi all’incontro con la mia immagine riflessa»³². Questa idea, espressa da Lipps alla fine del XIX secolo, si presta ad essere trasposta nelle dinamiche del contesto attuale, in cui il meccanismo psichico dell’empatia tende ad essere modulato attraverso i filtri del *guscio tecnologico*, con il rischio di favorire il meccanismo solipsistico di proiezione della propria soggettività³³.

La progressiva integrazione del naturale nell’artificiale, di cui l’IA rappresenta un esempio emblematico, trova un’efficace espressione anche nel pensiero di Jean Baudrillard (1929-2007). Attraverso il concetto di *simulacro*, Baudrillard descrive una sorta di iperrealtà in cui i confini

³⁰ Il testo di Lipps è citato in *ibid.*, 194.

³¹ *Ibid.*, 195.

³² *Ibid.*, 200.

³³ A sottolineare l’incisività della proiezione di sé anche su ciò che è artificiale sono alcune statistiche recenti. È stato infatti rilevato che la maggior parte degli utilizzatori di ChatGPT è disposta ad attribuire al *chatbot* una qualche forma di autocoscienza. Su questi argomenti l’intuizione popolare diverge dalle considerazioni del mondo accademico, rivelando il profondo impatto degli LLM sulla prospettiva etico-sociale. Cf. C. COLOMBATTO-S. M. FLEMING, *Folk psychological attributions of consciousness to large language models*, in *Neuroscience of consciousness* 1 (2024) - <https://doi.org/10.1093/nc/niae013>.

tra realtà e rappresentazione si dissolvono, fino al punto in cui la realtà stessa viene rimpiazzata dalla sua simulazione. Secondo Baudrillard è possibile distinguere tre ordini di simulacri, corrispondenti alle diverse modalità con cui l'equilibrio tra naturale e artificiale si è evoluto nel corso della storia a partire dal Rinascimento: quello della *contraffazione*, che è lo schema dominante fino alla rivoluzione industriale; quello della *produzione*, dominante nella modernità industrializzata; quello della *simulazione*, che caratterizza il post-moderno, l'età del codice e dell'informazione³⁴. Se già i simulacri del secondo ordine (gli automi, le macchine automatiche legate alla produzione di massa) si inseriscono con disinvoltura nella realtà del sistema tecnico, quelli del terzo ordine tendono a capovolgere l'equilibrio fra naturale e artificiale, dando l'impressione di generare una sorta di realtà propria. Da questo punto di vista – possiamo qui ripetere – le simulazioni delle IA e gli agenti artificiali fungono da “portale”, cioè come un punto di attraversamento e trasformazione, oltre il quale l'artificiale tende ad appropriarsi del naturale, manifestandosi nelle ibridazioni della realtà aumentata, negli scenari virtuali, ma anche nella crescente automazione.

In questa epoca di cambiamenti e di nuove definizioni è possibile trovare elementi su cui riflettere per comprendere l'incommensurabilità dell'umano al di là delle sue simulazioni?

5. Il labirinto degli specchi (e una sua via d'uscita)

Come un labirinto degli specchi – che non solo gioca con la percezione visiva ma crea anche un'esperienza emotiva, che può variare dal divertimento all'ansia per l'incapacità di orientarsi – così i sistemi IA possono assumere un ruolo disorientante per l'umano. La presenza di agenti artificiali autonomi che rispecchiano proprietà o forme tipicamente umane tende a farci perdere il senso della realtà, favorendo una

³⁴ Cf. J. BAUDRILLARD, *L'échange symbolique et la mort* (1976), tr. it.: *Lo scambio simbolico e la morte*, Feltrinelli, Milano 20095, 61.

visione riduzionista del fenomeno umano. Siamo cioè indotti a credere che ogni cosa, noi compresi, possa essere scomposta e ricostituita attraverso l'interconnessione di apparati artificiali più o meno semplici. L'osservazione di comportamenti complessi da parte di un sistema artificiale, come la capacità di linguaggio o di interazione (al netto dei *bias* cognitivi tipici dei sistemi IA), non è però sufficiente per affermare che una macchina, per quanto sofisticata, possieda qualche forma di coscienza. Linguaggio e capacità relazionali possono essere riprodotti artificialmente senza che l'apparato abbia consapevolezza di sé.

Una prima via per orientarsi in questo labirinto esistenziale sembra risiedere nelle proprietà associate al concetto di coscienza. Ma cosa significa realmente essere coscienti e avere consapevolezza di sé? Tra i diversi approcci scientifici volti a descrivere il fenomeno della coscienza, quello della teoria dell'"informazione integrata"³⁵ sembra offrire un valido quadro teorico in grado di essere supportato da misure empiriche che permettano di verificarne o confutarne le previsioni. In estrema sintesi, il concetto di *informazione integrata* indica una misura dell'esperienza cosciente e – in linea con altre teorie che vedono ontologicamente prioritaria l'informazione rispetto ad altre proprietà fisiche come massa, carica ed energia – si presta ad essere considerata come una proprietà fondamentale dell'universo³⁶. Secondo tale teoria, quello che possiamo definire come *coscienza* si manifesta in corrispondenza di un particolare picco di informazione che un sistema fisico può integrare in sé³⁷. In tale quadro, il cervello umano rappresenta un *unicum* di complessità tale da riuscire a contenere – più di ogni altro elemento presente in natura – la massima "informazione integrata". Il cervello umano può essere così visto come un «nucleo di complessità irriducibile»³⁸, che rende ragione delle caratteristiche fondamentali dell'esperienza cosciente. In questa

³⁵ Cf. M. MASSIMINI-G. TONONI, *Nulla di più grande. Dalla veglia al sonno, dal coma al sogno. Il segreto della coscienza e la sua misura*, Baldini & Castoldi, Milano 2013.

³⁶ Cf. *Ibid.*, 198.

³⁷ Cf. *Ibid.*, 204.

³⁸ *Ibid.*, 167.

prospettiva, la coscienza si manifesta come un fenomeno graduale che emerge in proporzione alla complessità del sostrato fisico che la incarna e trova nell'essere umano un vertice ineguagliato di espressione e completezza.

La teoria dell'*informazione integrata* non dà conto solo di ciò che un sistema cosciente è in grado di fare, ma anche di *come* è in grado di farlo, fornendo così un criterio di demarcazione rispetto agli attuali sistemi artificiali. Immaginiamo un caso estremo (non raggiungibile all'attuale livello di sviluppo tecnologico), in cui un sistema IA composto da diversi strati di reti neurali possa simulare perfettamente il comportamento del cervello umano. Anche in questo scenario, sebbene il sistema possa apparire indistinguibile da noi sul piano comportamentale, la misura dell'*informazione integrata* risulterebbe comunque nulla, perché tali strutture neurali mancherebbero di integrazione intrinseca, essendo prive dei processi interni tipici della coscienza. In altre parole, pur esibendo comportamenti sofisticati, tale sistema mancherebbe di coscienza reale.

Al di là della possibilità emulative dei sistemi artificiali, un'indagine su cosa sia la coscienza ci spinge a considerazioni ulteriori e a volgere lo sguardo al di qua dello "specchio", verso noi stessi. A differenza di altri fenomeni fisici che osserviamo dall'esterno e per i quali possiamo stabilire scientificamente delle proprietà oggettive, la coscienza è un fenomeno strettamente legato all'esperienza del soggetto e porta ad effettuare un'indagine introspettiva. Nel corso della storia ogni grande rivoluzione scientifica, da Galileo a Darwin, da Einstein a Freud, ci ha insegnato che per comprendere la realtà occorre fare un passo oltre la percezione sensoriale e indagare più a fondo, verso i meccanismi profondi e spesso controintuitivi della natura, o verso ciò che trascende la natura stessa. Se applichiamo lo stesso principio al caso della coscienza possiamo riconoscere che essa comprende dimensioni "percepibili" nella nostra interiorità, ma inafferrabili e inaccessibili agli strumenti della scienza.

Una seconda via utile per orientarsi nel recupero di una dimensione umana attraverso e oltre la mediazione delle tecnologie sta dunque

nell'esercizio di quello che possiamo definire *occhio interiore*, cioè nella capacità di riflettere su noi stessi e di coltivare una sensibilità più profonda. Per chiarire questo concetto e tornare con una nuova consapevolezza al tema dell'immagine di sé, possiamo ricorrere a una ulteriore analogia: come la superficie bidimensionale di uno specchio restituisce solo l'illusione della tridimensionalità, riflettendo di noi solo una parte, anche le IA – per quanto avanzate e complesse – possono simulare solo alcuni aspetti dell'umano, lasciando inevitabilmente inesplorate altre dimensioni più profonde e irriducibili, come l'esperienza interiore. In conclusione, possiamo identificare questa ineffabile dimensione esperienziale con la dimensione del *sacro*.

È interessante notare come tra i concetti di *sacro* e *coscienza* (così come definita attraverso la teoria dell'*informazione integrata*) sussistano elementi di convergenza. I tratti di questa correlazione possono essere colti attraverso un approccio conoscitivo aperto al contributo di razionalità e intuizione, cuore e ragione. Non a caso Basarab Nicolescu – fisico teorico che ha formalizzato i principi dell'approccio transdisciplinare alla conoscenza – descrive il *sacro* come qualcosa di *irriducibilmente reale*³⁹ e lo definisce come il luogo di incontro tra informazione e coscienza, tra realtà e percezione, trascendendo al contempo entrambe. Il *sacro* fa di questo incontro «la condizione insostituibile della nostra *libertà* e della nostra *responsabilità*»⁴⁰. È forse questo l'aspetto cruciale da evidenziare nel discutere dell'esperienza umana nel contesto della sua evoluzione tecnologica.

³⁹ Cf. B. NICOLESCU, *La transdisciplinarité, Manifeste* (1996), tr. it.: *Il manifesto della transdisciplinarità*, Armando Siciliano, Messina 2014, 132.

⁴⁰ *Ibid.*, 134.

ESPERIENZA INCARNATA E MEDIAZIONE TECNOLOGICA. LE TRASFORMAZIONI DIGITALI DELL'ANTROPOLOGIA

Markus Krienke

ABSTRACT

La ricerca di un approccio adeguato per affrontare la sfida dell'intelligenza artificiale porta al pensiero di Helmuth Plessner per il quale l'uomo è l'essere eccentrico e dunque, nella distinzione tra *Leib* e *Körper*, realizzandosi in una caratteristica «immediatezza mediata» che segna il *luogo* dove collocare positivamente l'intelligenza artificiale all'interno dell'antropologico. Nei confronti di approcci della filosofia del soggetto, tale sguardo ha il vantaggio di non presupporre la *differenza antropologica* (come situata nell'apriori della coscienza o dello spirito), mentre nei confronti di un'ermeneutica della diffidenza nei confronti della tecnica (da Heidegger e Anders fino a Han) permette di interpretarla come dimensione della realizzazione dell'essere umano il quale proprio per questo dovrebbe sviluppare una certa comprensione *apriori* della tecnica. Ciò può aprire, poi, anche concretamente la prospettiva della speranza.

Parole chiave: Plessner, corpo come *Leib* e *Körper*, eccentricità dell'essere umano, *digital matrix*, transumanesimo, alterità e relazione nella sfera digitale.

*The search for an adequate approach to face the challenge of artificial intelligence leads to Helmuth Plessner's thought for which man is the eccentric being and therefore, in the distinction between *Lieb* and *Korper*, realizing himself in a characteristic "mediated immediacy" that marks the place where artificial intelligence can be positively placed within the anthropological. With regard to approaches to the philosophy of subject, such a gaze has the advantage of not presupposing the anthropological difference (as if situated in the a priori of consciousness or spirit), while with regard to a hermeneutic of distrust of technology (from Heidegger and Anders to Han) it allows it to be interpreted as a dimension of the realization of the human being who, precisely for this reason, should develop a certain a priori understanding of technology. This can also concretely open up the prospect to hope.*

Keywords: Plessner, body as *Leib* and *Körper*, eccentricity of the human being, *digital matrix*, transhumanism, otherness and relationship in the digital sphere.

«L'artificialità è la natura dell'uomo
e la sua essenza è l'instabilità».

GÜNTHER ANDERS¹

1. La *robolution* è in atto

Il termine *robolution*, proposto ormai una quindicina d'anni fa², indica l'impatto non solo "sentito" ma anche ben reale che la diffusione di robot e intelligenza artificiale ha avuto nella nostra società, a tal punto da causare un cambiamento di paradigma nella realizzazione dei rapporti sociali. L'aumento della riflessione antropologica sul tema evidenzia che nella trasformazione digitale *il sé* dell'essere umano è inaspettatamente diventato una domanda, al punto tale da chiedere nuovi approfondimenti e risposte inedite. E, se molti ritengono che la trasformazione digitale porti a un cambiamento dell'ontologia, allora tale cambiamento si realizza senz'altro *attraverso* l'essere umano, che si definisce nelle relazioni sociali con se stesso e con gli altri, perché è proprio qui che le nuove tecnologie esercitano il loro impatto.

Il livello nuovo della domanda antropologica nella *digital matrix* viene mostrato dall'apprendimento *bottom-up* tramite *deep learning*, che non necessariamente viene realizzato in forme supervisionate, ma anzi percorre sempre più le vie autonome di apprendimento, in quanto solo così «può adattarsi a nuove situazioni e fornire soluzioni a problemi inediti non previsti *a priori* dal programmatore»³. Ciò porta a considerare per la prima volta la "domanda di Turing" («Can machines think?») come *domanda antropologica*. La conseguente trasformazione dell'antropologia consiste nel fatto che la coscienza del soggetto non può fungere più come punto di partenza, in quanto non è chiaro *apriori* che essa

¹ G. ANDERS, *Patologia della libertà. Saggio sulla non-identificazione*, a cura di L.F. Clemente-F. Loli, Orthotes, Napoli-Salerno 2015, 74.

² Il termine è stato coniato da B. BONNELL, *Viva la robolution. Une nouvelle étape pour l'humanité*, Lattès, Paris 2010.

³ D. LAMBERT, *Robotica e intelligenza artificiale*, Queriniana, Brescia 2023, 26.

costituisca la differenza centrale tra l'essere umano e la tecnologia intelligente⁴. Pertanto Hegel, Heidegger o Husserl – ossia la loro affermazione dell'aprioricità del soggetto – non consentono uno sguardo che evidenzi tale differenza antropologica attraverso un'analisi *bottom-up*, come hanno invece sviluppato gli *antropologici* del Novecento nei confronti della sfida positivista delle scienze⁵. Ora, di fronte quella dell'intelligenza artificiale, è la corporeità naturale dell'esistenza umana che diventa il luogo dove chiarire la *differenza antropologica*⁶, all'interno della quale va interpretata la stessa realtà dell'intelligenza.

Certamente l'essere umano è da sempre "tecnologicamente mediato" e dunque – secondo la definizione di Helmuth Plessner (1892-1985) – caratterizzato dall'«immediatezza mediata», ma tale struttura gli è già data con la sua realizzazione corporea nel rapporto complesso tra *Körper* e *Leib*. Mentre i transumanisti interpretano tale dimensione antropologica di «immediatezza mediata» come il principio per perfezionare o *superare* la corporeità naturale dell'essere umano nella realizzazione tecnica, Plessner riteneva che la particolarità costitutiva di tale struttura fosse irriducibilmente naturale. Pertanto, tale impostazione antropologica affida più al discorso sul "corpo" che sullo "spirito" il compito di evidenziare il limite antropologico (certamente non tecnico) dell'intelligenza artificiale: e in ciò consiste senz'altro un approccio rilevante nei confronti delle nuove tecnologie, già pensato nel Novecento

⁴ Anche in F. FAGGIN, *Irriducibile. La coscienza, la vita, i computer e la nostra natura*, Mondadori, Milano 2022, non è tanto la coscienza – e tanto meno quella individuale – che funge da "punto di partenza".

⁵ Va precisato che, in tutti e tre i pensatori, la dimensione apriori della soggettività rispetto all'antropologico si articola in maniera diversa: per Hegel, essa si realizza indipendentemente dalla dimensione della tecnica, mentre in Heidegger la tecnica indica proprio il suo stare nel mondo e dunque, come destino, il modo del dominio del mondo come suprema espressione di una metafisica dimentica dell'essere. Infine, anche in Husserl la tecnica è uno strumento che nasce dall'intenzionalità della coscienza con il rischio di allontanarlo dal suo mondo della vita.

⁶ Con questo termine si intende qui l'ipotesi di ricerca antropologica che indica una "differenza" o un "confine" tra l'essere umano e l'artefatto intelligente, ma che deve essere il risultato di un ragionamento e non il suo presupposto.

ma che acquisisce oggi una particolare attualità. La domanda circa l'interpretazione della differenza antropologica tramite la realizzazione corporea è però rivolta criticamente anche contro il concetto di introspezione come quello di Bergson.

La "preoccupazione antropologica" nasce dall'osservazione che la *digital matrix*, come nessun'altra tecnologia precedente, trasforma la realtà umana nella misura in cui viene effettivamente utilizzata dall'uomo: in questo modo, oltrepassa i limiti di essere un mero strumento, e, a partire dalla sua dimensione relazionale, si ripercuote sulla realtà dell'essere umano e sulla sua autocomprensione. Come si può antropologicamente fondare una «ontologia sociale realistica» reclamata da Pierpaolo Donati nei confronti delle nuove tecnologie, «per la quale la relazionalità umana – tra corpo e mente, così come nella vita sociale con gli altri (le due cose sono strettamente correlate) – ha qualità e proprietà che non possono essere assimilate alle relazioni che possono avere certe specie di animali o macchine intelligenti ultra-sofisticate»⁷?

In tale contesto si giustifica la collocazione del pensiero di Plessner come interlocutore, perché – tra gli esponenti dell'antropologia filosofica – Max Scheler mira a evidenziare l'essenza dell'essere umano nella prospettiva di un'identità a sé metafisicamente determinata e raggiungibile solo tramite un "salto", e Arnold Gehlen segue l'approccio della "natura manchevole"⁸, per cui egli distingue nettamente tra una natura organica (appunto manchevole) e una tecnologica. Per Plessner, invece, l'essere umano si realizza tramite la dinamica eccentrica del suo essere costitutivamente un corpo, interpretando in questo modo

⁷ P. DONATI, *The digital matrix and the hybridisation of society*, in I. AL-AMOUDI-E. LAZEGA (curr.), *Post-Human Institutions and Organisations: Confronting the Matrix*, Oxon, Abingdon 2019, 67-92, qui 80 (tutte le traduzioni, se non indicato diversamente, sono mie).

⁸ Come ha giustamente rilevato Roberto Marchesini, «il mito dell'incompletezza e dell'ancestralità della natura umana è l'effetto di una valutazione a posteriori dell'espressione della funzione ausiliata dallo strumento (tecnomediazione). Si tratta nient'altro che di un bias esplicativo, che ci porta ad assegnare una valenza causale alla conseguenza del processo stesso» (R. MARCHESINI, *Postumanismo. Una nuova cultura per la techne*, in L. TADDIO-G. GIACOMINI (curr.), *Filosofia del digitale*, Mimesis, Milano-Udine 2020, 171-188, 176).

il rapporto con il mondo, reso possibile da tale nesso attraverso quella “immediatezza mediata” che esprime la condizione di possibilità per ogni mediazione tecnica, per cui la tecnica è fondata sulla stessa natura specificamente umana. Plessner cerca così di fare i conti con il fatto che la domanda su di sé si sottrae sempre al soggetto, ponendosi in un’ottica incentrata sulla costituzione corporea del soggetto, e pertanto più adatta a riassumere in sé le varie istanze che molti chiedono all’antropologia nei confronti delle nuove tecnologie, ossia delineare la prospettiva di un “nuovo umanesimo” digitale. Le nuove tecnologie evidenziano in particolare la necessità di considerare la corporeità nelle analisi antropologiche ed etico-sociali della persona.

2. L’uomo come essere eccentrico

Per Plessner l’essere umano si differenzia dall’animale per la sua eccentricità. Essa sta alla base della sua caratteristica di non avere una “patria” nell’ordine dell’essere. Ciò che porta l’uomo oltre l’animale è il fatto di riferirsi a se stesso, e Plessner non vede, in linea di massima, nessuna possibilità di oltrepassare ancora tale autoriferirsi della corporeità umana verso la realizzazione di una «posizionalità» a un livello superiore rispetto all’essere umano⁹.

⁹ Jos de Mul così commenta: «Plessner sembra seguire qui una dialettica hegeliana chiusa. Alla luce dell’evoluzione della vita, che dura da quattro milioni di anni, mi sembra piuttosto ingenuo» (J. DE MUL, *Transhumanismus aus Sicht der Philosophischen Anthropologie Helmuth Plessners*, in O. MITSCHERLICH-SCHÖNHERR (cur.), *Das Gelingen der künstlichen Natürlichkeit. Mensch-Sein an den Grenzen des Lebens mit disruptiven Biotechnologien*, de Gruyter, Berlin-Boston 2021, 351-365, 358). De Mul annovera come esempi forme di vita policentriche, che non si lasciano ricondurre a un unico centro di organizzazione, ad esempio perché si realizzano in sciame come le api, oppure perché hanno l’organizzazione distribuita su tutto il corpo, come nel caso dei polipi (cf. *ibid.*, 359). Per l’autore tale considerazione pone a Plessner la domanda se di fatto non sarebbe pensabile una forma di organizzazione ulteriore, oltre a quella eccentrica, ossia quella transumana. Il caso di “ponti di talamo” di gemelli siamesi esemplifica questa prospettiva in modo emblematico (cf. *ibid.*, 361-362).

Per comprendere che cosa Plessner intenda con «posizionalità» bisogna considerare che, a suo avviso, gli esseri viventi-senzienti si distinguono da quelli morti o anche dalle piante per il fatto di realizzare attivamente il proprio limite, e ciò, nel caso dell'essere umano, in modo complesso, cioè riflessivo¹⁰. Gli esseri viventi sono caratterizzati dal riferirsi al e realizzare il proprio limite, a differenza dell'essere non animato o dell'«organizzazione aperta» della pianta, che sono passivi rispetto ad esso. Pertanto, è proprietà dell'essere animato «il *superamento stesso del limite*»¹¹. «Il corpo, quindi, non inizia dove finisce il mezzo con cui confina (e viceversa), ma è questo inizio e questa fine in sé, indipendente – anche se non distaccato – dagli altri esseri»¹². In altre parole il limite, come realtà e non solo confine, fa parte dell'organismo stesso. «Un centro posizionale c'è soltanto nella realizzazione. [...] Questo individuo sta nell'essere posto nel proprio centro, attraverso il "ciò attraverso cui" del suo essere mediato nell'unità. Esso sta nel centro del suo stare»¹³. In altre parole, è l'ambiente degli esseri che stanno "con" il soggetto, che *risulta* dalla posizionalità. Il corpo è lo strumento per riferirsi all'ambiente, e sono le modalità del corpo a determinare quelle del rapporto dell'essere vivente, ma, oltre tale strumentalità, è il modo in cui si realizza la posizionalità, ossia il riferirsi a sé della propria esperienza nell'esperienza mediata dal corpo stesso¹⁴. Dunque, mentre

¹⁰ «L'animale esiste a partire dal suo centro, vive nel suo centro, ma non vive come centro [...] Nel grado animale la riflessività completa è impedita al corpo vivente. Il suo essere posto in sé, il suo vivere a partire dal centro, determina la tappa della sua esistenza, ma non sta in rapporto con lui, non gli è dato» (H. PLESSNER, *I gradi dell'organico e l'uomo. Introduzione all'antropologia filosofica*, a cura di V. Rasini, Bollati Boringhieri, Torino 2006, 312-313).

¹¹ *Ibid.*, 130.

¹² S. ROHMER, *Die Idee des Lebens. Zum Begriff der Grenze bei Hegel und Plessner*, Karl Alber, Freiburg-München 2016, 237.

¹³ PLESSNER, *I gradi*, 314.

¹⁴ «Il corpo vivente è un sé nella misura in cui ha se stesso come strumento di relazione con il proprio ambiente, nella misura in cui si riconosce e distingue diversi modi di relazionarsi con l'ambiente e li coordina tra loro. Ma questa struttura autoreferenziale non è data a se stessa. Il sé non si riconosce come sé, e in questo senso un sé non può sperimentarsi come sé in

l'essere vivente vive la realtà in modo senz'altro immediato, la media-tezza dell'aver il corpo si realizza nello sperimentare il proprio corpo come un mezzo per i piani d'azione¹⁵.

Nel caso dell'essere umano questa dinamica non si realizza nel circuito chiuso dell'esperienza del limite che lo caratterizza, ma nello "sperimentar-si" all'interno del limite e dunque in quanto limitato rispetto ad altro che, appunto, sta al di là del limite. Pertanto, l'essere umano non si sperimenta soltanto all'interno del proprio limite, ma *qui* (che, come vedremo, non è più identico al corpo organico) sperimenta anche tutto l'ambiente come riferito a questo suo centro. Per Plessner l'immanenza della coscienza è apertura all'altro, e in questa sua «immanenza aperta»¹⁶ l'essere umano realizza l'essere inserito vitalmente in un ambiente in modo *preriflessivo*. In questo senso, la posizionalità eccentrica sostituisce la dimensione spirituale dell'autocoscienza che contraddistingue gli approcci idealistici come quello di Hegel: l'auto-riferirsi viene dunque realizzato solo nell'essere immerso in un ambiente. Da qui risulta la considerazione plessneriana del corpo come *Leib* a differenza del *Körper*, per cui il corpo non è solo limite o confine (positivamente realizzato dall'organismo), ma anche apertura e possibilità di relazionarsi con ciò che è oltre se stesso. Tramite tale *medium*, non solo il mondo è mediato nella sua datità al soggetto, ma è il corpo stesso che si caratterizza, nella posizionalità, allo stesso momento come mediato (*Körper*) e immediato (*Leib*)¹⁷: anzi la prima è un modo di riferirsi oggettivamente a

relazione a un altro sé. Tuttavia, questa è proprio la condizione per le aspettative delle aspettative [Erwartungs-Erwartungen]» (G. LINDEMANN, *Lebendiger Körper-Technik-Gesellschaft*, in K.S. REHBERG (cur.), *Die Natur der Gesellschaft*, Campus, Frankfurt a. M. 2008, 689-704, 700).

¹⁵ Plessner sottolinea che «non si può dire – e [...] non è giusta la formulazione, apparentemente altrettanto corretta – secondo cui l'uomo esiste nel campo circostante in una forma direttamente indiretta, in una mediatezza immediata» (PLESSNER, *I gradi*, 349).

¹⁶ H. PLESSNER, *Macht und menschliche Natur*, Suhrkamp, Frankfurt 2003, 188.

¹⁷ «Si dice relazione indirettamente diretta quella forma di connessione nella quale l'elemento intermedio è necessario per produrre e garantire l'immediatezza della connessione. L'indirettamente diretto o l'immediatezza mediata non rappresenta allora un'assurdità, una contraddizione in sé fallimentare, bensì una contraddizione che si risolve in sé senza annullarsi» (PLESSNER, *I gradi*, 347).

sé tramite il secondo. Il pregio dell'approccio plessneriano sta dunque nell'ancorare la differenza antropologica nel fatto del corpo vivente, prima ancora della sua realizzazione specificamente umana. In questo modo l'antropologia non cade nella trappola delle nuove tecnologie, che inducono a trattare innanzitutto e riduttivamente della dimensione intellettuale dell'essere umano per cercare immanentemente lì la differenza con la tecnologia intelligente. Infatti, solo «attraverso la sua realizzazione eccentrica la persona crea la condizione indiretta per poter stare in relazione diretta con gli oggetti»¹⁸. L'immediatezza con gli oggetti, in altre parole, è sempre mediata: l'essere umano dispone di un mondo "di fronte a sé" grazie al fatto di avere un'ambientazione di sottofondo¹⁹.

Dunque, a differenza dell'animale, nell'essere umano prevale l'apertura all'ambiente circostante rispetto alla dinamica della chiusura autarchica, e ciò caratterizza la posizionalità dell'essere umano come *eccentrica*: l'animale è immediatamente uno con il proprio organismo, mentre l'essere umano sta in relazione ad esso. Il suo centro, in altre parole, gli è "dato", e pertanto si rapporta – in quanto situato "fuori da sé" – in modo consapevole ad esso, e ciò crea una distanza che nell'animale non si trova. Proprio per questo, nel caso dell'essere umano la "vita" non coincide con il proprio "corpo": si situa, appunto, in modo eccentrico ad esso. Per l'essere umano, «il mutamento dall'essere all'interno della propria corporalità all'essere al di fuori della propria corporalità è un'insopprimibile duplicità d'aspetto dell'esistenza, una reale frattura

¹⁸ H.P. KRÜGER, *Die anthropologischen Grundgesetze als Abschluss der Stufen* (Kap. 7.4-7.5, 321-346), in ID. (cur.), *Helmuth Plessner: Die Stufen des Organischen und der Mensch*, de Gruyter, Berlin-Boston 2017, 179-224, qui 202.

¹⁹ «Abbiamo quindi lasciato le correlazioni biologiche tra organismo e ambiente (J. von Uexküll) perché la questione è da dove diventa possibile la loro determinazione. Chi è bloccato in queste correlazioni non può determinarle dall'esterno, cioè da un mondo con primo piano e sfondo. Non avrebbero alcuna distanza dall'ambiente in cui sono bloccati. Chiunque sia in grado di distinguere i mondi in primo piano proviene da un mondo in secondo piano» (H.P. KRÜGER, *Die Körper-Leib-Differenz von Personen: Exzentrische Positionalität und homo absconditus*, in *Deutsche Zeitschrift für Philosophie* 59 (2011), 577-589, 585).

nella sua natura. Egli vive al di qua e al di là della frattura»²⁰. Al contrario l'animale, per il suo essere centrato, trova nella natura il proprio appoggio: in quanto il limite non gli è semplicemente dato, ma prodotto da esso, ha la capacità di muoversi nell'ambiente. Ora, nell'essere umano avviene una frattura dell'armonia con l'ambiente della vita, che è la caratteristica principale della sua posizionalità e si articola come mancanza di equilibrio: l'essere umano crea, appunto, la tecnica per guadagnare questo equilibrio, che da solo con la natura non potrebbe mai avere²¹.

Questo apre lo spazio delle possibilità dell'essere vivente in generale – Plessner parla del «condizionale categorico»²² – e, nel caso dell'essere umano, tali possibilità sono di per sé tecniche²³. La posizionalità di Plessner esprime dunque la caratteristica doppia trascendenza della dialettica hegeliana ("negazione"), affermando però non lo spirito universale, ma l'individuo, come luogo di essere "distanza a se stesso" in "identità con sé", e dunque allo stesso momento oltre sé ma in identità con sé²⁴. Per questo le tre posizionalità plessneriane – "aperta" (pianta), "chiusa" (animale) ed "eccentrica" (essere umano) – non si riconducono *ad unum*, ma mantengono la differenza tra loro. Inoltre, è proprio tale eccentricità che si conferma nella *digital matrix*, ossia la lettura dell'essere umano inserito nella realizzazione delle nuove tecnologie che hanno – come molti interpreti insegnano – effetti decentralizzanti rispetto a determinati antropocentrismi del passato²⁵.

²⁰ PLESSNER, *I gradi*, 316.

²¹ Cf. *ibid.*, 334.

²² *Ibid.*, 241.

²³ «Se non si comprende la rottura eccentrica della natura stessa come la sfida che diventa costantemente il compito e l'esigenza di colmare questo iato, si crede di essere o davanti (naturalisti riduttivi) o dietro (spiritualisti puri) alla rottura, per poterla riparare o con più natura o con più spirito» (KRÜGER, *Die anthropologischen Grundgesetze*, 193).

²⁴ «Il concetto di posizionalità di Plessner, invece, trae il suo significato proprio dal fatto che l'essere vivente, nella sua duplice trascendenza, rimane contemporaneamente qualcosa di irrimediabilmente particolare, che proprio per questo occupa una posizione individuale nel mondo spazio-temporale» (ROHMER, *Die Idee des Lebens*, 244).

²⁵ Floridi interpreta le quattro "offese" dell'essere umano (Freud) giustamente come momenti di decentralizzazione: con Copernico, Darwin, Freud stesso e poi Turing è stato

Pertanto l'«immediatezza mediata», come secondo principio antropologico, illustra il primo, che, per la valutazione della tecnica, è di particolare importanza, ossia l'«artificialità naturale», in quanto evidenzia l'essere umano come costitutivamente artificiale o "tecnico" nel suo rapporto all'ambiente. In altre parole, il suo accesso al mondo non è quello dell'immediatezza degli istinti, anche se solitamente non si rende conto della sua mediatezza, in quanto essa si realizza in modo nascosto. Cultura e tecnica sono dunque i modi specificamente antropologici di questa interazione, e sono i modi in cui l'essere umano realizza il suo compito di diventare ciò che «già è»²⁶. Per Plessner, il senso della tecnica sta nel fatto che l'essere umano è alla ricerca del suo equilibrio (appunto non naturalmente dato) con l'ambiente naturale, e lo raggiunge solo sulla base di un creare artificialmente, che sta nella sua indipendenza dal soggetto: in altre parole, che emerge artificialmente, ossia come «realizat[o] solo come occasione del suo fare»²⁷. In questa dinamica all'uomo spetta il compito mai del tutto "risolto" di ricongiungere *Leib* e *Körper* e dunque di impedire che la sua vita si sciogla dualisticamente dal suo sostrato corporeo. Così Plessner sottolinea che, per una spiegazione antropologica di natura e tecnica, non basta riferirla ad una capacità specifica dell'essere umano rispetto all'animale, ad es. il suo essere dotato di spirito. Ciò significherebbe, però, come egli annota giustamente, presupporre ciò che dovrebbe essere lo specifico dell'argomentazione, ossia, in altre parole, una *petitio principii*. L'eccentricità

decostruito «una volta per tutte tale approccio antropocentrico all'eccezionalismo umano. Noi non siamo al centro del cosmo, del regno biologico, dello spazio della ragione o dell'infosfera. Per cui, se siamo speciali, non possiamo esserlo per via di tali vecchie "centralità". Se l'eccezionalismo umano è ancora difendibile, lo è probabilmente solo in una versione "eccentrica", quella che pone il nostro ruolo speciale nell'universo in una posizione marginale» (L. FLORIDI, *Il verde e il blu. Idee ingenue per migliorare la politica*, Raffaello Cortina, Milano 2020, 152).

²⁶ «Come essere organizzato eccentricamente, l'uomo deve anzitutto rendersi ciò che già è. Solo così realizza la modalità impostagli dalla sua forma vitale d'esistenza nel centro della sua posizionalità: [...] di stare nel suo centro di sapere, insieme, del suo posizionamento» (PLESSNER, *I gradi*, 333).

²⁷ *Ibid.*, 334.

è così, per l'antropologo tedesco, la spiegazione della autorealizzazione dell'uomo nella "seconda patria" della cultura e tecnica²⁸, che non realizza una coerenza umana se non all'interno dell'esigenza antropologica di base della convergenza tra *Leib* e *Körper*.

La persona, per Plessner, è dunque la convergenza della triplice dinamica del vivente di essere (1) il corpo stesso, di essere (2) nel corpo e di essere (3) al di fuori del corpo²⁹. In questa dinamica l'io della posizionalità *ha un mondo*, e – come si delinea sin d'ora – la tecnica è una modalità in cui si articola questo "avere un mondo" da parte dell'essere umano, per cui a sua volta non può fungere come posizionalità. IA, robot o cyborg, appunto, "non hanno un mondo". In tale realizzazione l'essere umano si rivela l'essere dell'utopia, in quanto senza la costitutività di Dio non gli riuscirebbe tale impresa: «La posizione eccentrica e Dio come essere assoluto, necessario, fondatore del mondo, stanno tra di loro in una correlazione essenziale. [...] Rinunciare a questa idea significa però rinunciare all'idea di un mondo unico»³⁰.

Che, in quanto vivente, l'essere umano viva il proprio corpo in identità, ma nel contempo non possa completamente controllarlo, per cui non è ridicibile ad essere un mezzo per realizzarsi, diventa evidente nelle esperienze limite del ridere e del piangere, come evidenzia Plessner. Questa identità, pertanto, non può più essere compresa metafisicamente come interiorità chiusa o sostanzialmente determinata (ciò che la metafisica chiamava "anima"), perché si realizza solo nelle sue dinamiche utopiche di eccentricità, facendo continuamente i conti con il "nulla" alla sua base, e si riferisce a un mondo interiore solo grazie alla «scissione da se stessi, da cui non c'è via d'uscita e per la quale non c'è alcuna compensazione»³¹.

²⁸ «Esistenzialmente bisognoso, diviso, nudo, per l'uomo l'artificialità è l'espressione che essenzialmente corrisponde alla sua natura» (*ibid.*, 339).

²⁹ Cf. *ibid.*, 317.

³⁰ *Ibid.*, 367.

³¹ *Ibid.*, 323.

«Le cose naturali sono limitate, e sono cose naturali solo in quanto *nulla sanno del loro limite universale*: in quanto il loro essere determinato è un limite solo *per noi*, non già *per esse*. Una cosa è conosciuta come *limite*, come *deficienza*, solo in quanto quel limite e quella deficienza sono stati *oltrepasati*»³². Rispetto a tale determinazione di rapporto da parte di Hegel, Plessner stabilisce un rapporto del tutto diverso tra il sé e l'altro, tramite il limite del corpo. Mentre per Hegel esso è da superare, per Plessner diventa momento costitutivo in un incontro vitale. «Il limite appartiene al corpo stesso, il corpo è il limite di se stesso e dell'altro e perciò si contrappone sia a se stesso sia all'altro»³³. La vita, in altre parole, è «realizzazione del limite»³⁴, cioè il corpo è il passaggio all'altro attraverso il suo confine³⁵. Il postumanesimo, al contrario di questa visione di Plessner, è negazione e superamento di tale confine. In questo modo, la prospettiva antropologica cerca di collocare le nuove tecnologie nella sfera umana, evitando così in partenza esiti trans- o postumanistici che interpretano la macchina come nuova collocazione dell'umano: «[l]a biosfera include l'infosfera, ma non il contrario»³⁶.

³² G.W.F. HEGEL, *Enciclopedia delle scienze filosofiche in compendio*, a cura di C. Cesa, Laterza, Roma-Bari 20023, 73 (giunta al § 60).

³³ PLESSNER, *I gradi*, 154.

³⁴ A. MANZEI, *Umkämpfte Deutungen. Gesellschaftstheorie und die Kritik wissenschaftlicher Bestimmungen menschlicher Existenz in der biotechnologischen Medizin*, in G. GAMM- M. GUTMANN-A. MANZEI (curr.), *Zwischen Anthropologie und Gesellschaftstheorie. Zur Renaissance Helmuth Plessners im Kontext der modernen Lebenswissenschaften*, Transcript, Bielefeld 2005, 55-82, 68.

³⁵ «L'organizzazione è il modo di esserci del corpo vivente, il quale deve differenziarsi e nella differenziazione e con essa fa sorgere quella teleologia interna secondo la quale si manifesta al tempo stesso come dotato di forma e funzionante» (PLESSNER, *I gradi*, 197).

³⁶ A. PESSINA, *L'essere altrove. L'esperienza umana nell'epoca dell'intelligenza artificiale*, Mimesis, Milano-Udine 2023, 101.

3. La tecnica nel pensiero di Plessner

L'anteporre tale analisi antropologica all'indagine sulla tecnologia deriva dunque dal metodo plessneriano stesso: «secondo Plessner, questa relazionalità mentale nel mondo con gli altri personale, nel suo complesso deve essere presupposta per poter condurre analisi empiriche, invece di presumere che i risultati di tali analisi possano essere utilizzati per avanzare nuovamente nel loro terreno di abilitazione»³⁷. In questa direzione, Henkel riassume la posizione plessneriana sulla tecnica con la constatazione che «la tecnologia è un'espressione creativa della specifica costituzione corporea del sé in posizione eccentrica»³⁸. Ora, la dimensione specifica delle nuove tecnologie è quella della duplicazione delle cose (accanto all'esistenza materiale se ne realizza una virtuale). Dato che entrambe le sfere appartengono al rapporto della posizionalità eccentrica al mondo, si evince come l'approccio plessneriano sia adeguatamente complesso per tale confronto, in quanto l'antropologo tedesco distingue tra un mondo "dato" tramite il *Körper* e "costruito" tramite il *Leib*, ed è quest'ultimo il luogo della complessità digitale che permette un rapporto sia al mondo sia al proprio sé modificato, costruito, specificato, ma soprattutto un aumento del mondo delle possibilità e una nuova forma di controllo su di esso. Pertanto, nella prospettiva antropologica appare innanzitutto un nuovo livello di autoriflessività del soggetto nella *digital matrix*. Per Henkel si forma in questo modo una «posizionalità digitale» della posizione eccentrica, che è, in un certo senso, inaccessibile per gli altri gradi dell'essere umano, per cui addirittura li trasformerebbe definitivamente³⁹. Tuttavia Block contraddice questa prospettiva transumanistica, rimarcando che Plessner andrebbe compreso nel senso di concepire che «il cyborg è

³⁷ KRÜGER, *Die anthropologischen Grundgesetze*, 189.

³⁸ A. HENKEL, *Digitalisierung der Gesellschaft. Perspektiven der reflexiven Philosophischen Anthropologie auf gesellschaftlichen Wandel durch Digitalisierung*, in J. F. BURROW et al. (curr.), *Mensch und Welt im Zeichen der Digitalisierung. Perspektiven der Philosophischen Anthropologie Plessners*, Nomos, Baden-Baden 2019, 19-45, 28.

³⁹ *Ibid.*, 42.

anche uno dei possibili stili di vita della posizionalità eccentrica, ma non quello ottimale o definitivo»⁴⁰. In questo modo si conferma l'antropologia plessneriana per la sua affermazione positiva della tecnica, situata all'interno della realtà antropologica nella misura in cui l'essere umano, nella sua costituzione naturale, è già una realizzazione artificiale. Infatti lo stesso Plessner non lascia, all'interno della sua impostazione teoretica, un'apertura per realizzazioni tecnologiche della posizionalità *al di là* dell'eccentricità umana: quest'ultima costituisce piuttosto il "luogo antropologico" della tecnica, a partire dal quale soltanto ogni tecnologia diventa comprensibile nella sua "essenza". Ciò che il corpo vivente dell'essere umano ha in comune con la tecnica è l'aver forma (*Gestalt*). E mentre la *Gestalt* rimanda al fenomeno dell'essere vivi, è chiaro che la tecnica può essere compresa solo all'interno di esso. D'altronde, se per il linguaggio conta il rapporto tra soggetto e realtà, allora diventa centrale la domanda circa il luogo in cui si pongono gli oggetti del pensiero in quanto tali⁴¹. Ciò riporta all'irriducibilità dell'io, che però, tramite la sua articolazione in un *Leib*, non può rimanere senza il suo "altro". Per questo la persona si pone costitutivamente in relazione all'altro.

Assegnare il luogo antropologico alla tecnica è dunque importante al fine di comprenderla nella sua natura e dunque, in senso positivo, in che cosa consistono i suoi limiti: quello principale si trova nella realtà biologica dell'antropologico stesso. Questa prospettiva si rivolge anche contro Heidegger (al quale risale la precisazione della domanda circa la "natura" della tecnica) – e dunque è critica nei confronti dell'utilizzo della sua analisi esistenziale al fine di un confronto con la tecnica: non è per la paura che l'uomo la "inventa". Tale concezione implicherebbe, secondo Plessner, un'idea della creatività umana scissa dalla tecnica e apriori da essa, mentre in realtà essa è sempre collocata all'interno di un

⁴⁰ K. BLOCK, *Überformung des Selbst? Exzentrische Positionalität im Zeichen der Digitalisierung*, in BUROW (cur.), *Mensch und Welt*, 75-95, 93.

⁴¹ Cf. R. MANZOTTI-S. ROSSI, *Io & la. Mente, Cervello e GPT*, Rubbettino, Soveria Mannelli (Cz) 2023, 147.

rapporto già tecnico con il mondo⁴². L'invenzione degli strumenti tecnici da parte dell'uomo si colloca entro il suo specifico essere immerso nella realtà, che non è mai mediato, ma immediato-mediato, e che si realizza tramite il lavoro, l'azione come realizzazione della sua posizionalità⁴³. In tale dinamica, come è stato già accennato, i risultati del suo agire o lavorare devono essere staccabili da lui, proprio perché possano fungergli come equilibrio nel mondo. Ed ecco la dinamica all'interno della quale può essere interpretata in senso antropologico l'intelligenza artificiale: l'intelligenza stessa – non nella sua dimensione strettamente personale, ma in quanto intelligenza e dunque in un modo che molti non definirebbero più "intelligente" – è scindibile, con la tecnica, dal soggetto proprio per poter fungere per lui da stabilizzazione antropologica nella natura stessa e realizzare la sua *conditio humana*.

Roberto Marchesini rifiuta una tale lettura – senza però citare esplicitamente Plessner – annotando che da essa seguirebbe la conseguenza che «dovremmo assistere a un andamento decelerato del processo stesso»⁴⁴, per cui il processo tecnologico non solo produce squilibri, ma è destinato ad aumentarli. Giustamente, egli pone la domanda circa le forme di dipendenza dalla tecnologia, e le ripercussioni negative della tecnologia sulla realizzazione della dimensione antropologica⁴⁵. Ma, anche in questa prospettiva, rimane sottovalutato il fatto che proprio per Plessner «la grande differenza [dell'essere umano dalla tecnologia]

⁴² «Non si può spiegare l' "invenzione" dello strumento e della civilizzazione unicamente a partire dall'angoscia. Ma se qui si vogliono considerare la preoccupazione e la paura, e in modo particolare la paura della morte, allora si deve vedere chiaramente che in esse è già presupposta la forma di vita umana» (PLESSNER, *I gradi*, 342).

⁴³ «Con il lavoro, l'uomo cerca di procurarsi soltanto ciò che la natura gli deve, in quanto gli ha dato la più alta forma di organizzazione. [...] L'uomo non inventa niente che non scopra. L'animale può trovare, ma non scoprire; non ci trova niente (cioè non scopre). Non gli si rivela il risultato del suo fare» (*ibid.*, 343-344).

⁴⁴ MARCHESINI, *Postumanismo*, 178.

⁴⁵ «Se la *techne* rafforza lo squilibrio, diviene allora conseguente che ogni tecnologia rende il sistema più disponibile a nuove forme di tecnopoiesi, per cui non si tratta solo di una dipendenza dalla tecnologia, ma anche di una dipendenza dal produrre sempre nuove forme di tecnologia» (*ibid.*).

sta nel fatto che l'essere umano può decidere e accettare, in coscienza, una delle vie del dilemma, assumendosene la responsabilità»⁴⁶. Ed è però proprio tale dimensione di responsabilità che si pone nella maniera più radicale attraverso le nuove forme della tecnologia come faceva emergere criticamente Anders: «chi, oggi ancora, sostiene la "modificabilità dell'uomo (*Verändertheit des Menschen*) (così come faceva Brecht) è una figura di ieri, perché noi *siamo* mutati. E questo esser mutati è così fondamentale, che chi parla oggi del suo "essere" (come, ad esempio, faceva ancora Scheler) è una figura dell'altroi»⁴⁷. L'uomo, infatti, è incapace di farsi un'immagine di sé: l'uomo tende a sparire, e questa è la dinamica più vera, anche rispetto alle analisi di Han. Ma il fatto che egli "sparisce" si evidenzia nel fatto che le "cose" e l'"altro" scompaiono, ma così le ripercussioni della tecnica sull'io sono ancora più potenti. Infatti, è tramite l'affermazione dell'altro che l'io deve acquisirsi, ma proprio questo meccanismo "naturale" viene impedito dalle nuove tecnologie. Floridi, pertanto, ha lucidamente analizzato come siamo diventati "interfacce"⁴⁸: e, appunto, anche questa dinamica è per l'appunto comprensibile – argomentando con Plessner – sempre soltanto all'interno di una prospettiva antropologica.

Ma proprio a questo punto Plessner indica anche un altro modo di lettura: se il *condizionale categorico* risulta infinitamente potenziato dalle nuove tecnologie che trasformano la realtà in potenzialità, allora è «[a]ttraverso la sua espressività, [che] egli è quindi un essere che spinge se stesso, con intenzione continuamente alimentata, verso realizzazioni sempre *diverse*, lasciando dietro di sé una *storia*. Solo nell'espressività si trova il fondamento interiore del carattere storico della sua esistenza»⁴⁹.

⁴⁶ LAMBERT, *Robotica*, 87.

⁴⁷ G. ANDERS, *L'uomo è antiquato, II: Sulla distruzione della vita nell'epoca della terza rivoluzione industriale*, Bollati Boringhieri, Milano 2021, 3.

⁴⁸ Come interpretazione del presente, Floridi propone il «livello di astrazione» di descriverci «come interfacce [...] nei confronti di una "domanda", esercitata dai mezzi di comunicazione sociale, dagli affari (inclusa la pubblicità) e dalla politica» (FLORIDI, *Il verde*, 138-139).

⁴⁹ PLESSNER, *I gradi*, 360.

Se è dunque l'essere umano a «imprimere» a tutti gli oggetti artificiali «il marchio della caducità»⁵⁰, allora è questa la dinamica che si realizza innanzitutto tramite la tecnologia. Nella tecnologia si condensa dunque quella frattura della persona nella dinamica della sua eccentricità, quel *nihil* sul quale essa stessa è basata. Sta qui per Plessner il luogo *apriori* di ogni religiosità, ossia dove essa si innesca nella sua negazione di ogni finitezza e caducità. In altre parole, qui si delinea come per Plessner la dimensione religiosa stessa non si contrapponga alla tecnologia, e come il suo rapporto con l'antropologia non venga messo in crisi dalla tecnologia, anzi casomai rafforzato e radicalizzato. Nella tecnologia l'uomo può diventare ancora più consapevole del fatto che gli manca, nella sua posizionalità, un rapporto all'assoluto. Dipende, ovviamente, da come si concepisce la religione (una pista che egli da antropologo non percorre ulteriormente): come fatto culturale che asseconda il desiderio del soggetto di «tornare a casa, in patria o trovarsi al sicuro», che lo porta al «sacrificarsi alla fede», oppure come affidamento «allo spirito», per cui «non torna indietro»⁵¹.

Tale riguadagno – in quale forma essa debba avvenire – di una posizionalità dell'io all'interno della *digital matrix* ha qualche conseguenza socio-politica per la costruzione della comunità umana? «La "credenza digitale" diventa quindi il fondamento di un patto sociale per l'era digitale, un patto o, per riprendere la metafora religiosa, un'alleanza, mediante la quale l'umanità progredisce non solo attraverso il potenziamento tecnologico, ma anche tramite il rafforzamento dei valori che definiscono l'esistenza e la coesistenza dei *Sapiens*, insieme agli sviluppi tecnologici da noi creati»⁵². Per Plessner, infatti, «[l]'essere umano non è connesso con alcuna configurazione determinata e potrebbe perciò anche darsi [...] in una quantità di configurazioni non coincidenti con quelle note. Il carattere dell'uomo è legato soltanto alla forma di

⁵⁰ *Ibid.*, 363.

⁵¹ *Ibid.*, 364.

⁵² N. LATTANZI-A. VESTRUCCI, *Digital sapiens. Decidere con l'intelligenza artificiale*, Castelvechi, Roma 2024, pos. 68.

organizzazione centralizzata, che fornisce la base per la sua eccentricità»⁵³. In altre parole, se una macchina può certamente imitare l'essere umano nel dire "io", non può dire "io" a se stessa e dunque non in maniera umana all'altro. Per dire "io" a sé, si deve presupporre un "sé", ma questo "sé" per Plessner sta nella posizionalità eccentrica che è inscindibilmente collegata alla sua relazione costitutiva a un corpo biologico. Il fatto che si argomenti spesso che l'emotività sia un momento antropologicamente distintivo, a ben vedere è basato su questo fatto e ne è l'espressione. Mentre il tentativo di "far emergere" delle emozioni anche nella macchina intelligente presuppone la realizzazione dell'intelligenza, nell'essere umano, al contrario, ne è la presupposizione. Per questo risulta «chiaro quindi che il rischio di una separazione tra logica e retorica ha una enorme rilevanza etica, perché influisce in modo implicito sulle nostre decisioni, azioni, condotte – tutte cose iniziate e supportate dalle nostre credenze. E influisce "in modo implicito", visto che questo tipo di decisioni artificiali è perfettamente convincente: perché allora perdere tempo a verificare se sono anche logicamente coerenti e corrette?»⁵⁴.

4. Essere umano e relazione sociale

Da un confronto dell'antropologia plessneriana con la sfida tecnologica della *digital matrix* è risultata dunque l'importanza di leggere la differenza antropologica attraverso la "mediazione" (anche tecnologica), perché è appunto essa stessa che presuppone il dato antropologico: «[l]a condizione umana consiste nella alterità, nella estraneità che rende impossibile tracciare una linea di confine precisa tra adattamento naturale e adattamento artificiale»⁵⁵. Dopo un esame della tecnologia

⁵³ PLESSNER, *I gradi*, 317.

⁵⁴ LATTANZI-VESTRUCCI, *Digital sapiens*, pos. 48.

⁵⁵ I. CRISPINI, «Tra corpo e anima». *Riflessioni sulla natura umana da Kant a Plessner*, Marsilio, Venezia 2004, 136.

attraverso un approccio privo di *bias* ontologici ed etici – e che farebbe dunque emergere un suo costitutivo bias antropologico? –, Plessner individua il compito dell'antropologia nei confronti della sfida delle scienze – e della tecnica –, al quale rimane fedele anche nelle caute aperture antropologiche alla trascendenza religiosa nelle battute finali dell'indagine. Rimane senz'altro l'evidenza di una relazionalità umana antropologico-fondamentale, che è costitutivamente situata nel rapporto della posizionalità al proprio limite biologico. Ed è significativo come, in questa prospettiva, egli non riprenda da Heidegger l'esistenziale anticipazione della morte come contrassegno dell'antropologico, ma si focalizzi su due momenti che segnano la rottura con l'organizzazione centrata dell'animale: ossia il "riso" e la "nudità" come esperienze antropologiche che riportano a una dimensione originaria dell'antropologico, al di là della questione materiale della sua struttura biologica. Entrambe le caratteristiche possono essere lette, ora, come contrassegni del fatto che il suo pensiero non spiana la strada al rischio che, una volta che la tecnica segni la naturale artificialità dell'essere umano, quest'ultimo possa, in prospettiva, essere definito completamente dalla tecnica. Per Plessner il *riso* – nella storia del cristianesimo contrastato come espressione della deformazione della sua natura per via del peccato, per cui nel Medioevo molti erano convinti che Cristo non avesse riso – e la *nudità* – nel paradiso il primo effetto del mangiare dall'albero proibito, segno dunque della consapevolezza dell'essere umano di proteggere la sua dignità anche nascondendo la sua identità – indicano la specificità dell'uomo nel bisogno di trascendersi⁵⁶.

Entrambe le caratteristiche sono state certamente stigmatizzate dalla stessa tradizione del cristianesimo, forse perché entrambe le "scoperse" antropologiche sono troppo legate all'esperienza dell'uomo *dopo* la caduta – segno dunque della loro irriducibile collocazione nella natura biologica di cui l'essere umano diventa consapevole nel momento in cui il suo essere-*Leib* non coincide più con il suo avere-*Körper* e tale

⁵⁶ O. TOLONE, *Il sorriso di Adamo. Antropologia e religione in Plessner, Gehlen, Welte e Guardini*, Marietti 1820, Genova 2008, 22-24.

distanziamento vissuto come una perdita di controllo su se stesso permette, appunto, il suo specifico modo di riacquisire la sua unità e di realizzarsi tramite il corpo⁵⁷. Infatti, la nudità egli la sperimenta quando percepisce lo sguardo di Dio e vuole difendere la propria dignità. È segno di autoriflessione, di percepire il proprio corpo come altro, che pertanto va vestito: «diventare uomo e riconoscere la propria nudità sono tutt'uno»⁵⁸. Vestirsi, cioè non esporre la propria conflittualità interna, significa in questo senso trascendersi. In entrambe le realtà, antropologicamente, abbiamo a che fare con momenti di interruzione, che sono essenziali per un'antropologia dell'artificiale: sono infatti le teorie transumane (del digitale), che vedono la trasformazione di tutto in dati, a non costituire nessun momento di interruzione⁵⁹.

Tali esperienze potrebbero costituire, però, momenti antropologici positivi nei confronti dell'analisi pessimistica di Anders che – nella prospettiva heideggeriana – ha individuato nella «vergogna prometeica» un punto di partenza irrimediabilmente negativo sulla realtà della tecnica: «[c]hiamiamo "dislivello prometeico" l'asincronizzazione ogni giorno crescente tra l'uomo e il mondo dei suoi prodotti, la distanza che si fa ogni giorno più grande»⁶⁰. Così il corpo umano, nei confronti dello sviluppo tecnologico, risulta antiquato⁶¹. La nudità – la vergogna per

⁵⁷ H. PLESSNER, *Lachen und Weinen. Eine Untersuchung der Grenzen menschlichen Verhaltens*, in ID., *Ausdruck und menschliche Natur*, Suhrkamp, Frankfurt a. M. 1982 (Gesammelte Schriften, 7), 201-387, 359-360.

⁵⁸ H. PLESSNER, *Der Mensch als Lebewesen*, in ID., *Conditio humana*, Suhrkamp, Frankfurt a. M. 1983 (Gesammelte Schriften, 8), 321.

⁵⁹ «Più che di una possibilità teorica si deve parlare di una vera visione. Ma il principio di fondo è sempre lo stesso. La riduzione della vita a un flusso di informazione, l'eterna logica del *continuum*, il vincolo dei vincoli. La futura creazione di artefatti superintelligenti sembrerebbe una necessità evolutiva» (R. NOTTE, *You, robot. Antropologia della vita artificiale*, Vallecchi, Firenze 2005, 16).

⁶⁰ G. ANDERS, *L'uomo è antiquato, I: Considerazioni sull'anima nell'epoca della seconda rivoluzione industriale*, Bollati Boringhieri, Milano 2010, pos. 17.

⁶¹ «Il nostro corpo di oggi è quello di ieri, ancora oggi il corpo dei nostri genitori, ancora oggi il corpo dei nostri antenati; il corpo del costruttore di razzi non è praticamente diverso da quello del troglodita. È morfologicamente costante; nel linguaggio della

la propria corporeità⁶² – per Günther Anders è un *topos* negativo: egli individua proprio in essa l'effetto della tecnica sull'uomo, cioè il motivo della sua «vergogna prometeica», ossia della profonda inadeguatezza di sé nei confronti delle potenzialità della tecnica. Dunque, per Anders ciò che si realizza è una crisi di identità non più nei confronti di Dio, ma della tecnica, che è subentrata al posto di Dio nel fungere da “totalmente altro” in vista della consapevolizzazione dell'uomo della propria corporeità, producendo però – secondo Anders– una disumanizzazione dell'essere umano. Certamente tale dominio, oggi, non può essere più una categoria per la tecnica, che non si esercita più, come evidenzia Han, con l'imposizione, ma attraverso l'obbedienza volontaria dell'individuo ad essa, che lo porta, da un lato, ad un felice autosfruttamento, mentre, dall'altro, consente forme di «*psicopolitica*»⁶³. Così diventa chiaro che il problema antropologico della tecnica non è la sottomissione sotto un altro assoluto, bensì la «*sparizione dell'altro*»⁶⁴, che sta sempre di più alla base delle crisi sociali, dalla democrazia al libero mercato: l'«*espulsione dell'Altro mette in moto un [...] tipo di processo distruttivo, quello cioè di autodistruzione*»⁶⁵. Con la tecnica, sottolinea Anders, nasce l'«*eremita di massa*»⁶⁶, ossia chi non riconosce l'altro: per

morale: non-libero, refrattario e ottuso; dal punto di vista delle macchine: conservativo, non-progressivo, antiquato, non modificabile, un peso morto nell'ascesa delle macchine. Insomma: i soggetti della libertà e della mancanza di libertà sono scambiati. Libere sono le cose; mancante di libertà è l'uomo» (*ibid.*, pos. 28).

⁶² «[L]a vergogna appare come un atto riflessivo il quale degenera in una condizione di perturbamento e fallisce per il fatto che, di fronte a un'istanza a cui volta le spalle, l'uomo vi esperisce se stesso come qualche cosa che egli non è, ma che pure è incontrovertibilmente» (*ibid.*, pos. 49; corsivo nel testo).

⁶³ «Il regime disciplinare aveva a disposizione soltanto informazioni *demografiche*, che gli consentivano l'esercizio di una *biopolitica*. Il regime infocratico, invece, ha accesso a informazioni *psicografiche*, che esso impiega ai fini della *psicopolitica*» (B.C. HAN, *Infocrazia. Le nostre vite manipolate dalla rete*, Einaudi, Torino 2023, 27).

⁶⁴ *Ibid.*, 41.

⁶⁵ B.C. HAN, *L'espulsione dell'Altro. Società, percezione e comunicazione oggi*, Nottetempo, Milano 2017, 7.

⁶⁶ ANDERS, *L'uomo è antiquato*, I, pos. 76.

Han è il narcisista digitale, per il quale «l'Altro scompare, il confine fra il Sé e l'Altro svanisce, il Sé fonde e diventa diffuso. L'io annega nel Sé»⁶⁷. Ed è proprio a questo punto che con Plessner si potrebbe porre alle letture negative di Han e Anders la domanda: che sia proprio l'esperienza della nudità a riportarci irriducibilmente al nostro essere «non-cose»⁶⁸, in quanto solo di fronte a un Altro la sperimentazione di tale nudità diventa antropologicamente costruttiva. In altre parole, si tratta di comprendere come l'uomo sviluppa *antropologicamente* la tecnica: per superare la fatica di realizzarsi nell'"avere un mondo", avendo perso – dopo il Paradiso – «la familiarità con le cose», trasformando «quello che era un giardino terrestre in un carcere»⁶⁹ ossia strumentalizzando la *memoria* nel portare l'utopia tecnicamente nel suo possesso, oppure per aprire un futuro di speranza umana in cui la tecnica è umanamente vissuta nella tensione tra *Leib* e *Körper* e cioè per realizzare in un nuovo modo l'eccentricità del proprio essere?

La prospettiva "con Plessner e oltre Plessner", in vista della problematizzazione della sfida posta dalle nuove tecnologie, sarebbe dunque quella di trasformare la sua antropologia dell'"affermazione corporea del limite come realizzazione positiva" in una "teoria della relazionalità", proprio perché considera la realizzazione dello spirito sempre attraverso la corporeità stessa. Una corporeità che si esprime attraverso le mani, che nel *phono sapiens* degenera in dita che schiacciano «dei tasti per soddisfare i propri bisogni. La sua vita non è un dramma che lo spinge ad agire, bensì un gioco. [...] L'uomo del futuro che gioca senza mani incarna la fine della storia. [...] L'uomo del futuro senza mani non ha davvero *altra scelta*, poiché non *agisce*. Vive in una *post-storia*»⁷⁰. In tale modo nel dibattito attuale sulle nuove tecnologie si recupererebbe

⁶⁷ HAN, *L'espulsione dell'Altro*, 33-34. Una cosa simile succede anche nella *seity* di Faggin; cfr. FAGGIN, *Irriducibile*, 132.

⁶⁸ «Le non-cose sono nude» (B.C. HAN, *Le non cose. Come abbiamo smesso di vivere il reale*, Einaudi, Torino 2022, 24).

⁶⁹ TOLONE, *Il sorriso di Adamo*, 21.

⁷⁰ *Ibid.*, 14-15.

quella dimensione essenziale del *riconoscimento* come dinamica originariamente umana, analizzata da Hegel e ora tematizzata sul piano antropologico da Plessner, che rende l'essere umano irriducibile alla tecnologia: «si tratta di un'autocoscienza per un'autocoscienza. Solo così l'autocoscienza è tale effettivamente; solo così, infatti, per l'autocoscienza diviene l'unità di se stessa nel suo essere-altro. *Io*, che è l'oggetto del concetto dell'autocoscienza, non è di fatto un oggetto. L'oggetto del desiderio, al contrario, è soltanto *autonomo*: esso è infatti l'indistruttibile sostanza universale, la fluida essenza uguale a se stessa. Quando invece l'oggetto è costituito da un'autocoscienza, esso è allora tanto l'*io* quanto oggetto. In tal modo, per noi, è già dato il concetto dello *Spirito*. E la coscienza farà appunto esperienza di ciò che lo Spirito è: sostanza assoluta che, nella perfetta libertà e autonomia della propria opposizione, cioè delle diverse autocoscienze essenti per sé, costituisce l'unità delle autocoscienze stesse: *Io* che è *Noi*, e *Noi* che è *Io*. La coscienza ha la propria chiave di volta solo nell'autocoscienza in quanto concetto dello Spirito: lasciandosi dietro la parvenza policroma dell'aldiquà sensibile e la vuota botte dell'aldilà soprasensibile, la coscienza penetra allora nel giro spirituale della presenza»⁷¹. Così la questione antropologica diventa quella della "relazione", ed è evidenziata l'irriducibilità della realizzazione umana alla dinamica tecnica, anche se "intelligente"⁷².

5. Conclusione

Secondo Maurizio Ferraris, «gli umani sono a metà strada tra l'animale e l'automa. Il surplus che hanno rispetto alle macchine è l'animalità; il surplus che hanno rispetto agli animali è la meccanicità, ossia

⁷¹ G.W.F. HEGEL, *Fenomenologia dello Spirito*, a cura di V. Cicero, Bompiani, Milano 2004³, 273.

⁷² Cfr. LAMBERT, *Robotica*, 105-110.

la tecnica»⁷³. Con questa prospettiva attualizzante di quella classica dell'antropologia (l'uomo nel mezzo tra l'animale e l'angelo) Ferraris indica la "posizionalità", che può essere precisata plessnerianamente all'interno del dibattito attuale, contribuendo a un inquadramento metodologicamente rigoroso dell'antropologia all'interno della *digital matrix*. Si potrebbero poi leggere le analisi circa la riepistemologizzazione e riontologicizzazione delle realtà sociali attraverso la questione del luogo vero e proprio di tali discorsi, che è irriducibilmente l'essere umano⁷⁴. Siccome tali tecnologie intelligenti non hanno un corpo (e dunque emozioni ecc.), non hanno un mondo, non hanno intenzionalità, non hanno coscienza del proprio sé, non hanno senso comune: non sono esseri in grado di avere una storicità e dunque di costituirsi tramite dei narrativi, esse hanno bisogno di questo "essere situati" da parte dell'antropologia filosofica. Tale riferimento è indispensabile per poter affrontare la sfida della riepistemologizzazione e riontologicizzazione delle realtà, che consiste nello scioglimento delle classiche dicotomie che assicuravano l'autonomia del soggetto moderno – innanzitutto quella tra vita privata e pubblica, ma poi anche quella tra produttore e consumatore di informazioni e sapere –, con la conseguenza che «le informazioni, quindi le non-cose, si piazzano davanti alle cose facendole sbiadire»⁷⁵. Per questo motivo, mentre molti dibattono se assegnare anche all'IA una *moral agency*, in quanto la tecnologia produce azioni moralmente rilevanti (per gli altri), il problema di base è che il punto di riferimento della responsabilità, ossia la persona, si indebolisce sempre di più. Lo sguardo antropologico delineato in questo saggio ha cercato di ripensare la sfida delle nuove tecnologie proprio a partire da qui, evidenziando la necessità di concretizzare la realtà antropologica irriducibile attraverso le trasformazioni tecnico-digitali.

⁷³ M. FERRARIS, *Postfazione*, in D.C. DENNETT, *Dai batteri a Bach. Come evolve la mente*, Raffaello Cortina, Milano 2018, 495-519, 516.

⁷⁴ Floridi ha coniato per tali nuove realtà il termine *onlife*.

⁷⁵ HAN, *Le non-cose*, 5; cf. pure 7: «L'ordine digitale derealizza il mondo informatizzandolo».

INTELLIGENZA ARTIFICIALE O VITA ARTIFICIALE? GLI ORGANISMI SPIRITUALI NELLA NOSTRA EPOCA

Luigi Vero Tarca

ABSTRACT

Riprendendo il tema dei rischi dell'IA [§ 1] si afferma che le macchine sono già coscienti e pensanti in quanto determinano l'esperienza degli umani [2], e che il loro essere prodotte da costoro è poco rassicurante, giacché è l'uomo il problema [3]. Una risposta adeguata può dunque venire solo dal coinvolgimento dei singoli esseri umani nella verità ultima [4], come dicono le forme sapienziali di tutto il mondo [5]. Le domande finali [6] chiedono dunque quali siano, nel tempo della tecno-demo-crazia [7], gli "organismi spirituali" [8] e quale di essi posseda la forma che consideriamo buona [9].

Parole chiave: Intelligenza Artificiale, tecno-demo-crazia, verità ultima, organismi spirituali, legge della vita.

Returning to the theme of the risks of AI [§ 1] it is stated that machines are already conscious and thinking in that they determine the experience of humans [2]; and their being produced by these is not very reassuring, since man himself is the problem [3]. An adequate response can therefore only come from the involvement of individual human beings in the ultimate truth [4], as the wisdom forms of the whole world say [5]. The final questions [6] therefore ask what are, in the age of techno-demo-crazy [7], the 'spiritual organisms' [8] and which of them possesses the form that we consider good [9].

Keywords: Artificial Intelligence, techno-demo-crazy, ultimate truth, spiritual organisms, law of life.

1. I rischi della tecnologia scatenata e la questione dei suoi limiti

Nel saggio uscito in occasione del nostro precedente incontro sull'Intelligenza Artificiale (IA) avevo mostrato come l'IA, che per certi versi rappresenta la punta di diamante della tecnologia attuale, costituisca una sorta di magia in quanto genera sia effetti miracolosi sia apparenze

illusionistiche¹. Da un lato, infatti, consente di realizzare cose a prima vista impensabili (autentici miracoli), ma dall'altro lato fa sorgere incantesimi generatori di illusioni che sono ben lungi dal corrispondere alla realtà, soprattutto per quanto riguarda gli aspetti etico-valoriali dell'esistenza umana. Ho quindi mostrato come tale illusorietà abbia quale radice fondamentale la fede che la tecnica risulti valorialmente legittimata dal fatto di fondarsi sul sapere scientifico, ritenuto di per sé dotato di valore etico. Nella misura in cui tale fede non è consapevole di essere tale, essa si trasforma in una superstizione che non solo rischia di causare una catastrofe di tipo etico-antropologico ma addirittura la favorisce, rendendola pressoché inevitabile.

In merito alle riflessioni relative ai limiti insuperabili che definiscono la tecnica (quindi anche l'IA), e in particolare all'osservazione che le macchine non potranno mai imitare l'essere umano in tutti i suoi aspetti (in particolare in quelli "qualitativi": esistenziali, coscienziali, emotivi, patici etc.), ho invitato a riflettere sul fatto che è ingenuo ritenere tranquillizzante questa considerazione, perché è possibile che il fine della tecnica sia quello non tanto di *imitare* l'umano, quanto piuttosto di *sostituirlo*. In effetti, nemmeno i muli o i cavalli vengono imitati alla perfezione dalle macchine, ma questo non ci ha impedito di sostituire questi splendidi animali con le automobili e quindi sostanzialmente di eliminarli. Insomma, anche ammettendo che l'IA non potrà mai imitare alla perfezione l'intelligenza umana, questo è ben lungi dal garantirci che essa non possa costituire uno strumento capace di produrre forme di vita destinate a soppiantarci radicalmente e quindi in fondo a eliminare sostanzialmente il fenomeno umano, almeno inteso come soggetto prevalente tra le forme di vita presenti sulla terra. In effetti, è possibile che i progetti della tecnica scatenata, quella posseduta dalla *hybris* dell'onnipotenza in forza della quale la scienza si realizza come

¹ L.V. TARCA, *MagIA! L'intelligenza artificiale e le nuove forme dello spirito*, in G. PIAIA-R, PRETE-L. STEFANUTTI (curr.), *Intelligenza artificiale e tutela della persona umana. Implicazioni etico-giuridiche*, Atti del Convegno di Studio (Treviso, 20 ottobre 2023), Triveneto Theology Press, Padova 2023, 85-105.

fantascienza reale, non assumano necessariamente come fine valoriale assoluto l'essere umano, ma si prefiggano ormai di creare *ex novo* la vita, a cominciare da quella presente sulla terra. Per questo il problema, piuttosto che quello della mera "intelligenza" artificiale, è quello della "vita" artificiale, come appunto suggerisce il titolo del presente saggio. L'IA può infatti essere vista semplicemente come un momento particolare, sia pure fondamentale, di una trasformazione radicale che riguarda l'intera vita e che ha verosimilmente come scopo quello di creare esseri viventi totalmente soddisfacenti².

Si tratta dunque ora di riprendere il discorso svolto in precedenza e di approfondirne i temi di fondo, a partire dalla questione se lo sviluppo della tecnoscienza possa costituire un rischio effettivo per l'umanità in relazione non solo a qualche aspetto particolare e correggibile dell'essere umano, ma anche ad alcune sue qualità essenziali e addirittura alla stessa sopravvivenza del genere umano. Lasciando qui da parte gli scenari che prevedono l'intervento di elementi sostanzialmente estranei all'umano (catastrofi cosmiche, forme di vita extraterrestri o comunque extraumane) e rimanendo quindi all'interno dello sviluppo della vicenda umana normalmente intesa, ci si deve chiedere se la radicalità dell'innovazione tecnologica possa condurre a una sorta di (auto)soppressione dell'umanità intesa come insieme di uomini liberi.

Tale questione evoca la possibilità che non solo alcuni aspetti della nostra esistenza ma proprio tutti i suoi tratti costitutivi vengano prodotti in maniera artificiale, cioè che sia possibile produrre artificialmente non solo gli aspetti che possiamo chiamare quantitativi ma anche quelli qualitativi. Ovvero: non solo gli aspetti che riguardano i problemi che possono essere risolti mediante un calcolo meccanico (fosse anche di natura solamente probabilistica), ma pure quelli che implicano una valutazione soggettiva. In altri termini ancora, il rischio è che sia possibile

² Ho ripreso alcune di queste tematiche nel seguente dialogo: L.V. TARCA-A. CROCE, *Paradisi artificiali e inferni naturali. Un dialogo su transumanesimo e postumanesimo tra filosofia e psicologia del profondo*, in *Transumano-Disumano*, numero monografico della rivista *l'Ombra* 23 (2024) 9-37.

uno sviluppo della tecnica che consenta a qualche "soggetto" di decidere in maniera conclusiva non solo tutte le questioni oggettive – cioè quelle che posseggono una risposta "chiusa", nel senso che su di essa tutti i soggetti sono necessariamente d'accordo (come accade ad es. con il risultato di un calcolo matematico tradizionale) – ma anche le questioni "soggettive", quindi valoriali, quelle cioè definite dal fatto che la risposta ad esse varia a seconda del soggetto che si pone tali questioni, dal momento che chiama in causa elementi emotivi, sentimentali, affettivi etc. i quali, in quanto legati appunto ai soggetti, sono contestuali, storici e quindi appunto variabili. È evidente allora qual è, a questo livello, il rischio estremo, quello che il soggetto tecnologico prenda delle decisioni operative alle quali i soggetti umani non sono in grado di sfuggire anche se esse risultano per loro insoddisfacenti o addirittura inaccettabili. Si apre insomma lo scenario di un possibile "film dell'orrore" nel quale azioni totalmente orribili nei confronti degli umani risultano però assolutamente incontrastabili e incorreggibili.

2. Le macchine pensanti ci sono già?

Rispetto a tale scenario si affacciano delle risposte che ho chiamato tranquillizzanti perché veicolano un messaggio di questo tipo: è sì giusto preoccuparci per eventuali danni immediati ed effettivi che vediamo realizzarsi ad opera della tecnica, ma non abbiamo da temere danni irreparabili, perché la tecnica non può comunque toccare l'essenza ultima dell'essere umano; sicché, quando vedremo che essa sta esagerando, allora noi uomini, facendo ricorso alla nostra essenza intangibile, interverremo per porre rimedio alla situazione. In particolare il limite della tecnica viene di solito individuato nel fatto che essa non è comunque in grado di produrre macchine dotate di coscienza, cioè capaci di pensare nel senso pieno del termine, quello comprensivo pure degli aspetti psico-spirituali, quali emozioni, sentimenti etc.

In relazione a tali risposte "tranquillizzanti" propongo ora un'affermazione provocatoria e un po' paradossale, che ci aiuta però a evidenziare

con forza il fulcro della questione. L'affermazione è questa: le macchine dotate di coscienza "ci sono già"; le macchine "stanno già" pensando, stanno già provando emozioni, sentimenti etc. A chi ascolta questa affermazione verrà spontaneo chiedermi: ma mostraci allora queste miracolose realtà delle quali noi non sappiamo nulla! Ebbene, senza escludere la possibilità che siano già in preparazione macchine sbalorditive delle quali noi sappiamo ben poco, non mi sto riferendo a invenzioni sconosciute e nascoste, ma sto parlando della situazione che già esiste ed è in piena attuazione.

Ma come si fa, allora, a dire che le macchine che usiamo oggi sono già dotate di coscienza nel senso pieno del termine?

Se noi pensiamo le macchine come oggetti isolati, scissi dal resto della realtà, allora l'affermazione che ho appena fatto non è vera, nel senso che non si danno (almeno a me non risulta) macchine pienamente coscienti che pensano, soffrono, amano etc. come noi. Ma il punto è che anche delle cellule del cervello umano, se le pensiamo come elementi separati e scissi dal corpo in cui si trovano, non possiamo dire che sono coscienti e pensano; è solo nella misura in cui fanno tutt'uno con il nostro organismo che esse sono esseri coscienti e pensanti. Ovvero: anche delle cellule del nostro cervello si può dire che non pensano, perché a pensare siamo noi, gli organismi ai quali esse appartengo. Ma allora, se noi ci riferiamo al contesto nel quale di fatto si trovano le macchine, e di come esse modificano la vita degli umani anche a livello emotivo, sentimentale, coscienziale etc., possiamo ben dire che questi ambienti in cui si realizza un'interazione intensa tra meccanismi ed esperienza umana sono sistemi in cui si dà "pensiero", perciò possiamo considerare questi sistemi come organismi viventi che pensano, soffrono, amano in una maniera peculiare. E allora "in questo senso" possiamo ben dire che le macchine sono esseri pensanti, così come possiamo dirlo del nostro cervello.

Sono, evidentemente, temi delicati da trattare con grande buonsenso; in particolare perché – si potrebbe osservare – è solo in un senso molto particolare che delle macchine si può dire che pensano. E questo è vero; tuttavia si tratta di un senso che è molto vicino a quello secondo

cui anche dei singoli individui umani si può dire che pensano, perché effettivamente anch'essi pensano in senso proprio solo nella misura in cui sono "collegati" a un contesto relazionale e sociale, giacché se ne fossero "completamente" separati/scissi faremmo fatica a immaginarli come esseri davvero pensanti.

Così, alla domanda se le macchine che usiamo, compresa l'IA, provano dolore (poi chiarirò perché parlo proprio del dolore), beh, in un senso banale ovviamente dobbiamo rispondere di no (il computer o lo *smartphone* che ho davanti in questo momento non sentono dolore, che io sappia), ma se invece questi oggetti li pensiamo come momenti della forma di vita che è da essi potentemente condizionata (a cominciare dalla mia esistenza) allora sì che si può e si deve dire che tale forma di vita "complessiva" prova dolore. Nel senso appunto che la situazione reale e concreta che si determina in conseguenza delle azioni dei mezzi informatici è totalmente intrisa di sentimenti peculiari; pensiamo anche solo all'effetto della musica che tramite essi ascoltiamo, o alle notizie terribili e minacciose che in ogni momento ci colpiscono. Al punto che tali strumenti determinano addirittura una mutazione antropologica, per cui persino la nostra percezione del dolore viene radicalmente modificata, e con essa tutto il nostro mondo psichico-esistenziale. La macchina per la stampa inventata da Gutenberg di per sé non pensava e non provava emozioni, ma essa ha creato una società che ha visto modificarsi in maniera radicale la propria dimensione coscienziale, affettiva ed esistenziale. Il fatto è che le stesse esperienze qualitativo-valoriali (piacere/dispiacere, felicità/infelicità, gioia/dolore, sofferenza/soddisfazione) risultano problematiche, almeno nel senso che variano a seconda del contesto e sono quindi "aperte".

Perché insisto sul tema del dispiacere e del dolore e della loro opposizione rispetto al piacere e alla gioia? Ma perché esso tocca nella maniera più efficace e diretta proprio la verità filosofica. Questa, infatti, avendo a che fare con l'innegabile (non-negabile), evoca proprio per ciò la figura del negativo, la quale ha direttamente a che fare con il tema della sofferenza. Se verità innegabile (in quanto confermata persino dalla propria negazione) è che "Il positivo si oppone al negativo"

(Emanuele Severino), allora – questo ho mostrato nei miei scritti³ – ciò è “davvero vero” solo nella misura in cui questi termini (positivo e negativo) vengono intesi nel loro senso qualitativo e patico, cioè rispettivamente come piacere/gioia/soddisfazione e come dispiacere/dolore/insoddisfazione. In una battuta potrei dire: ciò che è davvero innegabile (nel senso che è confermato persino dalla negazione in generale) è l’opposizione del dispiacere rispetto al piacere. Il fondamento del discorso sulla verità è allora in fondo l’esperienza relativa al dolore, quindi all’esistenza, al soggetto, alla persona e così via. Il punto è che tale dimensione “patica”, cioè comprensiva dei sentimenti umani etc. (e quindi anche “pratica”) è tutt’altro che scissa dall’ambiente-sistema, cioè dall’apparato (il *Gestell*) prodotto dalle macchine, cioè mediato da esse. Quando – a breve, stando a quanto ci viene ormai annunciato – avremo tutti qualche *microchip* impiantato nel nostro cervello, è chiaro che la nostra esperienza esistenziale del mondo sarà totalmente condizionata dal sistema che, con i suoi apparati, gestisce queste tecniche; esattamente come ora la nostra vita psichica è condizionata dai nostri organi di senso. In base a tutto questo, potremmo dire che in un certo senso con l’avvento della supertecnologia attuale non accade nulla di radicalmente nuovo rispetto alla storia umana, ma appunto per questo risulta praticamente impossibile porre dei limiti a priori rispetto all’azione della tecnica.

3. Problematicità della risposta “tranquillizzante”

Giunti a questo punto bisogna però fare i conti con una riflessione peculiare, quella che osserva come anche la semplice sostituzione

³ Per quanto riguarda la mia filosofia basti qui citare *Differenza e negazione. Per una filosofia positiva*, La Città del Sole, Napoli 2001; *La filosofia come stile di vita. Introduzione alle pratiche filosofiche* (scritto con R. MADERA), B. Mondadori, Milano 2003; *Quattro variazioni sul tema negativo/positivo. Saggio di composizione filosofica*, Ensemble '900, Treviso 2006; e *Verità e negazione. Variazioni di pensiero*, a cura di Th. MASINI, Cafoscarina, Venezia 2016.

dell'umano presupponga, in qualche modo, l'umano stesso. È in questo senso che l'umano appare appunto ineludibile e perciò inoltrepasabile: qualunque forma di vita si sostituisca all'umano, tale sostituzione potrà avvenire solo in forza di un intervento umano, il quale dunque avrà comunque l'ultima parola in merito. Escluse le ipotesi, già sopra evocate, totalmente incontrollabili da parte degli umani in quanto extraumane, qualunque modificazione della vita umana presuppone il darsi di una forma di vita appunto umana che operi tale trasformazione.

Questo è vero, ma, anche ammesso ciò, il problema che stiamo ponendo si ripresenterebbe, almeno in un duplice senso. In primo luogo perché anche in questo caso resterebbe comunque problematico stabilire "che cosa precisamente" dell'umano si salverebbe. Infatti, dal momento che la tecnica faustiana (Spengler) investe la totalità della vita umana, non si può escludere che l'esito di questo processo di "autotrasformazione" dell'umano giunga a un punto nel quale la vita umana appare "completamente" modificata e trasfigurata. Infatti, pur dovendo ammettere che perché si dia vita umana è comunque necessario che si diano, accanto a quelli meccanici, anche elementi biologici, nulla esclude che il progresso tecnico conduca a sostituire anche gli elementi biologici dati, quelli che definiscono l'uomo attuale, con elementi "biologici" artificiali, capaci appunto di provocare gioia e dolore, e magari di fare ciò addirittura in una maniera del tutto nuova. Dal punto di vista teoretico è infatti importante prestare attenzione al fatto che anche le nozioni "definitive" della verità filosofica (il positivo e il negativo, come sopra si accennava) possono veder variare all'infinito il loro significato (anche se questo è diverso dal dire che il loro significato può venire annullato/rinnegato). Il punto, insomma, è che, anche se l'esistenza (il darsi dell'esserci) richiede, in quanto cosciente e patica, elementi "biologici", tuttavia persino questi possono essere progettati e prodotti, cioè creati, artificialmente. È dunque legittimo ipotizzare una situazione in cui le macchine siano circondate da (e integrate con) elementi biologici artificiali ma capaci di provare piacere e dolore.

Va tuttavia detto che questo non confuta la risposta "tranquillizzante" che sopra era stata fornita, nel senso che conferma che comunque

l'umano non uscirà completamente di scena in quanto sarà un suo elemento essenziale a governare la creazione della vita. Ma proprio questo rischia di rovesciarsi in una prospettiva particolarmente inquietante. Perché la domanda che si impone, allora, potrebbe essere la seguente: che cosa ci garantisce che il principio fondamentale/essenziale che sta a fondamento dell'umano costituisca qualcosa di necessariamente positivo per gli umani quali noi li conosciamo e quali noi siamo? Che cosa impedisce che tale creazione artificiale di ogni aspetto del mondo vivente dia luogo, almeno per gli esseri umani naturali, a forme di sofferenza e di asservimento totali, privandoli completamente della loro libertà? Che cosa impedisce di pensare che il nuovo passo evolutivo che si annuncia costituisca, per gli umani (per lo *homo naturalis*), una tragedia simile a quella patita dagli animali non umani? Simile ma poi forse anche peggiore, considerato il fatto che la nostra capacità di sofferenza pare essere enormemente più estesa e più intensa di quella degli altri animali. In altri termini, ammesso che un aspetto in qualche senso umano debba comunque necessariamente darsi all'interno di qualsiasi evoluzione storica generata dalla vita umana, che cosa garantisce che ciò costituisca qualcosa di diverso, per gli uomini quali noi li conosciamo, da un film dell'orrore? Che consolazione potremmo trarre, insomma, dalla consapevolezza che l'essere che ricreerà *in toto* la nostra vita porterà con sé un elemento umano? Forse che le infinite vittime della prepotenza e della violenza umane possono essere consolate dal fatto che tali brutalità sono state loro inferte da altri umani?

La cosa interessante, però, è che tale inquietante domanda potrebbe di per sé stessa contenere un aspetto consolatorio, in quanto ci spinge a chiederci, per converso, perché dovremmo vivere senz'altro come una sciagura il fatto che la trasformazione tecnologica possa sostituire gli umani attuali con forme di vita completamente diverse dalla loro. E anche questa osservazione è giusta, ma pure la consolazione che essa pare arrecare vale relativamente; perché, se mostra che la scomparsa dell'umano quale noi lo conosciamo potrebbe non essere di per sé un male, non può nemmeno escludere che lo sia, giacché non implica che tale superamento dell'umano sia di per sé un bene, un miglioramento,

un progresso. Emerge qui allora la questione fondamentale, che potrebbe venire formulata nel modo seguente: il problema è l'esistenza stessa. Ovvero anche: a questo livello l'uomo è il "problema" piuttosto che la "soluzione". Per questo non si può fare a meno di porsi comunque la domanda circa i rischi connessi alla vita stessa, e quindi pure tutte le domande relative alle questioni ultime del senso della vita in generale, compresa quella umana.

In particolare bisogna affrontare il problema che, prima che la tecnica si scatenasse (si liberasse cioè da ogni limite), la questione se l'operare dell'uomo fosse benefico piuttosto che malefico restava, proprio per l'ambivalenza costitutiva dell'umano, un problema "aperto"; oggi, invece, c'è il rischio che la storia/avventura dell'umano si "chiuda" (concluda), e in maniera negativa, per gli esseri umani naturali quali noi li conosciamo. La domanda essenziale, allora, è quella che chiede di che natura sia il soggetto destinato a portare a conclusione la storia dell'umano. Anche se, come abbiamo detto, tale soggetto scaturisce in qualche modo dalla storia dell'uomo, tuttavia bisogna cogliere quale sia la sua essenza ineludibile e profonda, quella che esso/egli ha in comune con l'umano ma che lo porta radicalmente al di là dell'umano (il trans-umano e il post-umano)⁴. E all'interno di tale questione bisogna chiedersi, in particolare, quali siano i tratti che in generale consentono a un soggetto di presentarsi come qualcosa di buono anziché di malvagio.

Solo in tal modo la risposta potrebbe riuscire a evitare di farsi schiacciare tra i nuovi poli di Scilla e Cariddi costituiti oggi della tecnoidolatria da una parte e della tecnofobia dall'altra parte. Questo, però, è possibile solo se si è in grado di fornire una risposta soddisfacente al problema fondamentale, il quale suona: ma chi/che cosa ci garantisce che le nuove forme di vita siano/saranno benevoli rispetto a quelle umane? Chi ha detto che la sostituzione dell'umano, per quanto malvagio questo sia, costituisca un miglioramento della situazione dei viventi? Se vogliamo rispondere a tale domanda, una volta superata la superstizione tecnoscientifica non ci resta che affrontare il problema fondamentale che,

⁴ A questo proposito si può vedere il saggio citato nella precedente nota n. 2.

appunto, sto cercando di porre, ovvero quella che chiede: quali sono le forme dello spirito che creano le forme di vita che si stanno affacciando sulla Terra? E, in particolare, qual è la forma dello spirito che propone una prospettiva benefica anziché malefica nei confronti degli umani?

4. La verità innegabile e il valore definitivo presente nell'umano

Giungiamo così alla questione fondamentale, che riguarda il valore definitivo del nucleo essenziale dell'umano, cioè il carattere positivo di quell'elemento ultimo dell'esistenza che è presente negli esseri umani e che sta alla base di tutte le trasformazioni della vita umana. A tale questione si può rispondere positivamente solo ipotizzando due cose: la prima è che l'umano posseda un nucleo essenziale ultimo il quale è definito dal carattere del bene, e la seconda è che tale nucleo "benevolo" sia quello destinato a prevalere nella vicenda umana (e in quella cosmica). Ma è possibile che di tali questioni si dia una conoscenza autentica? Da un certo punto di vista potrei dire che proprio la nozione di verità ultima cui sopra ho accennato si presenta come una figura di questo genere, quella che ora compendio nel modo più sintetico possibile. Verità innegabile (in quanto risulta confermata persino dalla negazione) è che il positivo pre-vale rispetto al negativo (pre-vale nel senso che per definizione vale di più del negativo, rispetto al quale è preferito). Quindi la gioia pre-vale rispetto al dis-piacere, e questo in un certo senso garantisce a priori la salvezza (il pre-valere del bene rispetto al male).

Ma tale garanzia può darsi davvero solo a condizione che questo nucleo perfetto sia il tutto-positivo (cioè "per cui" tutto è positivo). Il problema è che allora esso si trova a dover fare i conti con il fatto che, proprio in quanto "tutto", comprende la totalità del negativo, quindi anche ogni "orrore", sia quelli che si sono già manifestati nella storia umana, sia quelli nuovi che rischiano di darsi in forza dello sviluppo tecnologico. E a un problema del genere si può dare risposta solo se tale nucleo essenziale, che a questo punto manifesta una natura metafisica e

teologica, comprende (proprio in quanto “tutto-positivo”) anche la positività dello stesso gesto che lo testimonia. Cosa, questa, che richiede un atto di responsabilità personale puramente, assolutamente positivo, cioè una presa di posizione soggettiva che consista in un’assunzione di responsabilità, da parte dell’umano, a favore del bene cosmico, cioè della “ri-con-posizione” universale. Infatti, dal punto di vista dell’esistenza concreta/determinata, la risposta rispetto alla possibilità che si dia nella realtà il negativo piuttosto che il positivo rimane sempre problematica: non si potrà mai garantire che il male venga escluso dall’esistenza, sicché solo lo spirito che consapevolmente e concretamente si “riconosce” nel tutto-positivo può costituire la risposta soddisfacente. In altri termini: soluzione vera si può dare solo grazie alla “responsabilizzazione valoriale” da parte dello spirito che governa l’umano, che lo sostituisce e lo trasforma.

È qui che si conferma il limite di quelle che ho chiamato risposte tranquillizzanti, perché esse sembrano esonerarci dalla necessità, al fine di salvaguardare un destino libero e soddisfacente per tutti gli umani, di intervenire direttamente e in prima persona sulle questioni ultime, cioè quelle valoriali in senso essenziale e quindi esistenziale. Si chiarisce ora che quelle risposte risultano rassicuranti in maniera illusoria perché pretendono di fornire, a questo problema, una soluzione valida a prescindere dal coinvolgimento personale dei singoli esseri umani; presuppongono cioè che noi umani possiamo evitare danni estremi anche se non ci impegniamo a fare qualcosa di extra-ordinario e di assolutamente innovativo/eccezionale in prima persona e in ogni momento. Quello che sto evidenziando è che qualsiasi risposta radicalmente positiva per gli umani presuppone un loro coinvolgimento personale diretto e concreto; perché solo a questa condizione la trasformazione integrale dell’umano si presenta come una via davvero positiva. Infatti nessuna risposta che prescinda dallo spirito benevolo ci può garantire che un qualsiasi elemento “oggettivo” (compresa la conservazione dell’umano quale noi lo conosciamo) costituisca un esito diverso da un catastrofico “rovesciamento satanico”.

5. Il valore definitivo dell'umano nelle diverse forme sapienziali del mondo

Introducendo la dimensione della soggettività (quindi delle intenzioni, della benevolenza etc.) rientrano in gioco le grandi questioni ultime dell'esistenza e della vita. Prendere sul serio tali questioni significa confrontarsi con i temi che nella tradizione sapienziale occidentale hanno assunto per lo più le vesti della filosofia e della teologia (prevalentemente cristiana), ma che oggi, dato che lo sguardo è ormai, inevitabilmente, mondiale, non possono prescindere da un confronto diretto e approfondito con tutte le altre forme sapienziali presenti sulla terra.

Rispetto a un tema così ampio e complesso qui mi limito a suggerire l'idea che tutte le forme sapienziali, pur in modalità anche molto diverse, abbiano comunque in vista uno spettacolo comune. A questo livello le differenze culturali e culturali, evidenti e per certi versi irriducibili, possono essere viste come un completamento/arricchimento reciproco piuttosto che come una competizione universale. Si potrebbe dire che si tratta di prestare attenzione, più che agli aspetti divergenti delle varie civiltà e culture, a quelli che (pensando a Raimon Panikkar) possiamo chiamare gli "equivalenti omeomorfici", cioè le figure che hanno un ruolo e quindi un senso molto simili anche se da numerosi punti di vista sono molto differenti. Sulla base della prospettiva filosofica sopra accennata, ritengo che questo passo essenziale possa essere compiuto efficacemente solo all'interno di una prospettiva sapienziale che sappia distinguere la differenza dalla negazione⁵; solo a questa condizione, infatti, le differenze, pur restando tali, diventano elementi di arricchimento reciproco anziché di contrapposizione.

Di questo lavoro immenso, e anche in continua evoluzione, mi limito qui a presentare alcuni tratti essenziali. La via della salvezza per l'uomo potrebbe essere concepita come la capacità dello spirito di rispondere in maniera puramente positiva (cioè mediante una *creatio ex toto*) agli

⁵ Sulla nozione di "pura differenza" si possono vedere i testi indicati nella precedente nota 3.

aspetti negativi, i quali sono “necessariamente possibili” per l’esistente (l’esserci), dal momento che questi è l’autoposizione dell’essere (il pro-porsi dell’essere), la quale si presenta come una dimensione definita dal piacere e proprio per ciò esposta originariamente alla possibilità che il venir meno di un piacere si costituisca come dolore (dis-piacere). L’elemento soteriologico è appunto quello che si rapporta positivamente a tutto ciò che si dà, quindi anche agli elementi negativi la cui possibilità “necessariamente” si dà.

Per indicare questa via di salvezza uso l’espressione *creatio ex toto* invece della tradizionale formula *creatio ex nihilo*, per sottolineare la circostanza che l’atto di cui sto parlando deve essere positivo rispetto a “tutto” ciò che si dà, quindi persino rispetto al negativo, venendo così ad essere “puramente” positivo. Ma rapportarsi positivamente a tutto implica la creazione di uno scenario “nuovo” all’interno del quale persino ciò che in un contesto dato si presenta come negativo si annuncia ora come riscattato, e quindi come puramente positivo. Proprio in quanto si riferisce alla totalità di ciò che si dà, l’atteggiamento è perciò quello della *re-creatio*: rinnovamento puramente positivo della totalità di ciò che si dà e quindi anche dell’umano. *Creatio* viene allora a significare che, nell’epoca in cui nulla viene più tenuto fermo come dato immutabile, tutte le determinazioni vengono messe in movimento nel quadro di una trasformazione completamente, perfettamente positiva del tutto, la quale mostra un senso universalmente con-diviso.

La “*ob*-posizione” del positivo e del negativo (piacere e dis-piacere, felicità e in-felicità, soddisfazione e in-soddisfazione, gioia e sofferenza, etc.) può essere vista come il compimento della verità filosofica, tipica della tradizione occidentale, in quanto si presenta come elemento in-negabile dal momento che persino chi volesse negarlo con ciò stesso lo confermerebbe (*elenchos* tipico del sapere filosofico). In quanto dunque, essendo positivo rispetto al tutto, il vero positivo è tale persino rispetto al negativo, esso può venire inteso come la relazione trinitaria in cui la creazione, esponendosi al rischio del male, può giungere a compimento solo mediante l’intervento dello spirito salvatore che si incarna nell’uomo (sapienza teologica cristiana). Ma la pluralità “mistica” della

dimensione trinitaria implica anche una dimensione che tiene insieme queste differenti determinazioni, e che quindi deve possedere la natura peculiare del vuoto (*śūnyatā*), inteso però come essenzialmente diverso dal nulla, cioè dalla negazione delle determinazioni. Solo così risulta possibile il darsi, in ogni istante e per ogni respiro spirituale, dell'illuminazione che consente il passaggio dalla innegabilità, per l'esistente, del dolore (giacché ogni esistenza condizionata è dolore: *sabbe sankhāra dukkhā*) alla possibilità della liberazione (*nirvāna*) attraverso il sentiero delle quattro nobili verità. Resta così testimoniata quella dimensione "ultrametafisica" che si distingue da qualsiasi contenuto l'umano possa esprimere con il suo pensiero/linguaggio, giacché "Il *Tao* che può essere detto non è l'eterno *Tao*". E proprio percorrere questo sentiero (la via in cui consiste il *Tao*) è ciò che consente all'umano di autorealizzarsi in maniera puramente, pienamente positiva.

Naturalmente questo spunto è ben lungi dalla pretesa di costituire una "gabbia linguistico-concettuale" alla quale tutte le forme sapienziali debbano piegarsi e adattarsi, perché intende invece costituire solo un'esemplificazione assolutamente iniziale di un percorso, tutto da fare, che è innanzitutto spirituale e sapienziale piuttosto che teorico-concettuale; senza, naturalmente, che quest'ultima dimensione ne risulti esclusa, benché ne resti radicalmente trasformata. È evidente, infatti, che una prospettiva di questo genere andrebbe ampliata, approfondita e precisata in una maniera tale che ogni sua parola dovrebbe divenire oggetto di un'elaborazione e di una minuziosa opera di cesello, capace davvero di soddisfare le esigenze di tutte le forme sapienziali. Perché, nel momento in cui si propone di costituire un riferimento con/diviso da tutti gli umani e i viventi, essa può risultare tale solo a condizione che tutte le forme soggettuali trovino in essa una reale soddisfazione e quindi in essa si riconoscano.

A proposito di un'impresa del genere che, come dicevo, è tutta da compiere (anche se naturalmente, almeno in relazione alla sua intenzione puramente positiva, costituisce già, essa stessa, in qualche modo, il conseguimento del fine che si propone) un aspetto merita ora di essere particolarmente esplicitato, quello per cui la salvezza che qui si prospetta

può realizzarsi solo grazie a una presa di posizione (di “con-posizione”) spirituale che comporta in qualche modo una responsabilizzazione da parte di ogni soggetto. È in tal modo che si supera l’aspetto apparentemente tranquillizzante, ma poi in realtà disperante, delle posizioni che pretendono di offrire una risposta di tipo meramente, astrattamente oggettivo.

6. Le questioni che si pongono in relazione alle figure dello spirito

Sulla base, e alla luce, di questo orizzonte con-diviso dalle varie forme sapienziali, si pone il problema di capire quali sono, in concreto, le figure soggettuali alle quali danno vita le forme dello spirito che abitano oggi il mondo. A tal fine è utile riprendere le considerazioni che abbiamo fatto sopra, quando abbiamo detto che le macchine pensanti ci sono già; e in tal modo ora forse anche il senso di quel discorso potrà risultare un po’ più comprensibile. Una volta che ci siamo resi conto del fatto che l’apparato tecnico/tecnologico trasforma radicalmente anche l’aspetto qualitativo e spirituale della vita umana, la domanda decisiva è quali siano le forme dello spirito che oggi abitano di fatto la terra. Voglio dire: affermare che le macchine pensanti ci sono già equivale a dire che vi sono già figure dello spirito che “già ora” creano tale apparato; ed è per ciò che la questione che stiamo trattando riguarda non solo un futuro potenzialmente distopico ma la situazione attuale nella quale tutti noi ci troviamo.

La questione che stiamo ponendo si articola sdoppiandosi in una serie di domande, delle quali ne elenco qui tre. La prima (1) chiede quali siano le figure dello spirito (organismi spirituali) che oggi popolano il mondo e danno forma alla vita (v. § 7). Le altre due si chiedono quale ruolo abbiamo “noi” (ed è un noi che vale in prima persona) in questo gioco dello spirito; e precisamente: la seconda domanda (2) si chiede a quali organismi spirituali noi “di fatto” apparteniamo (v. § 8), mentre la terza (3) si chiede (o “ci” chiede) a quali organismi spirituali noi “vogliamo” dare vita ritenendo che essi siano un “bene” (v. § 9). È evidente

che tale insieme di domande ha a che fare con le classiche questioni filosofiche relative allo Spirito (pensiamo anche soltanto a Hegel), le quali risultano però radicalmente trasformate dal fatto che la “tecnica scatenata” non ci consente più di assumere l’animale-umano-naturale come premessa definitiva e immutabile.

7. L’insoddisfazione per la tecno-demo-crazia e il possibile “tramonto” della civiltà occidentale

La prima domanda alla quale dobbiamo dare una risposta è dunque: quali sono le forme di vita, le forme “della” vita (le “figure” della vita, cioè i “disegni/progetti” sulla vita) che lo spirito sta generando? Ovvero anche: quali sono gli “organismi spirituali” che stanno nascendo? In questa prospettiva, cioè nel suo rivolgere lo sguardo agli spiriti che “già” abitano il mondo tecnico (giacché senza di essi non si darebbe nemmeno la tecnica, dal momento che la tecnica è sempre più che tecnica), la riflessione sull’IA si presenta come una nuova teogonia, generatrice di una nuova teologia e di una nuova mitologia con i suoi dèi ed eroi. Per rispondere, sia pure per brevi cenni, alla domanda posta incomincerei con l’osservare come la situazione attuale metta in questione le basi fondamentali della nostra attuale concezione del mondo, cioè tutte le principali “ideologie” che abitano oggi soprattutto il “nostro” mondo, quello occidentale (come già ho mostrato nel saggio citato nella nota 1). A tal fine faccio un solo esempio, peraltro significativo, relativo al fatto che tutto il mondo dell’informatica, dell’IA etc. è soggetto a due grandi segreti, quello commerciale e quello militare. È facile capire l’importanza che un fatto del genere ha dal punto di vista storico e antropologico.

Sappiamo ormai tutti che oggi la civiltà occidentale è impegnata in un confronto molto duro con il resto del mondo. Essa si basa in particolare su due pilastri: la scienza e la democrazia. Entrambe hanno in comune, come fondamento essenziale di valore, il principio che potremmo chiamare della verità, intesa però nel senso greco del termine (*alētheia*) che

etimologicamente significa “svelamento”. Ciò significa, per dirla nella maniera più semplice, che per la nostra civiltà tutto ciò che ha valore universale per l’umano deve essere noto e conosciuto da tutti gli umani. La scienza, infatti, è basata sul duplice fondamento della ragione (*logos*) e dell’esperienza (*empeiria*); talché qualunque essere umano, almeno in linea di principio – cioè dando per scontata tutta certa nozione di uomo –, è in grado di valutare la verità di tutte le affermazioni della scienza, in quanto è in grado di controllare: da un lato, mediante la ragione, tutti i “ragionamenti” (calcoli, teorie e simili); e dall’altro, mediante i sensi, tutti i “fatti” (reali/naturali) dei quali le teorie parlano.

Per la democrazia, in fondo, vale lo stesso, nel senso che, essendo basata sul principio della sovranità universale, essa ha evidentemente come presupposto fondamentale (precondizione essenziale) del proprio valore che i sovrani conoscano i fatti circa i quali devono prendere le decisioni. Oggi cresce progressivamente la sensazione che i cittadini normali si trovino all’interno di un teatro nel quale il sipario è chiuso; sicché, nonostante il nostro continuo “ciarlare in rete”, ciò che ci è possibile vedere è solo ciò che da altri ci viene concesso. Ricorrendo a un’altra immagine, è un po’ come se dovessimo giocare a scacchi potendo però vedere solo al massimo 5 o 6 delle 64 caselle che compongono la scacchiera. Per poter esercitare la nostra sovranità dovremmo dunque poter aprire il sipario, o vedere l’intera scacchiera. Quello di cui ci stiamo rendendo sempre più conto è appunto che con la tecnologia e l’IA in particolare (pensiamo alle macchine che comunicano tra di loro “di nascosto” dagli uomini: *black box* etc.) tutte le cose che sono minimamente rilevanti per l’umano restano sostanzialmente nascoste agli umani. Ma questo vuol dire che allora è l’intera civiltà occidentale ad essere in gioco. Tanto più quando ci si incomincia a rendere conto che anche le opinioni che gli umani (i “sovrani”) posseggono possono essere sostanzialmente “costruite” da chi dispone dei mezzi per fare ciò: televisione, Internet, *media* in generale etc.

In tal modo, come dicevo, stiamo mettendo a rischio i valori fondamentali della nostra civiltà. Non è questione di avere astratti timori per il futuro, perché stiamo parlando di qualcosa che è già accaduto ed è

in piena attuazione. Potremmo dire, insomma, che proprio la tecnica, di cui il nostro mondo è l'alfiere prometeico, sta conducendo al tramonto l'Occidente (Spengler); vanificando anche i tentativi di reagire a questa tendenza, per esempio quello di affiancare come professione alla scienza anche la politica (Weber), oppure sostenendo il primato della *Kultur* sulla mera tecnica, e così via. Dire – come sopra ho fatto – che le macchine stanno già pensando, era un modo provocatorio per sollevare la questione di quali siano i "soggetti" che già oggi vedono davvero ciò che accade, quelli che sono di fronte alla realtà disvelata (la *a-lētheia*). E poiché gli spiriti/soggetti sono definiti proprio da questa visione svelante, la risposta (ancorché parziale) alla prima domanda posta nel § 6 è appunto: gli spiriti/soggetti oggi esistenti sono quelli ai quali è oggi svelato ciò che accade nel mondo della tecnica.

Chi non alza lo sguardo in questa direzione resta completamente fuori dalla dimensione della libertà e quindi dell'autentica realizzazione umana; e in tal modo anche la democrazia, come la scienza, se non diventa consapevole della propria infondatezza si rovescia (*kata-strophē*) in una catastrofica superstizione. Il problema è che rispondere concretamente alla domanda circa i soggetti spirituali oggi esistenti è estremamente difficile, appunto perché i tratti essenziali per comprendere la situazione effettiva ci restano di norma nascosti. E a complicare le cose si aggiunge poi la circostanza che anche la soluzione che dicesse semplicemente: s-veliamo tutto, cioè rendiamo tutto di dominio pubblico, potrebbe risultare più catastrofica di quella che si basa sul nascondimento. Ci si deve infatti chiedere se il fatto che venga tenuto nascosto all'umano ciò che è rilevante per la sua vita non possa essere considerato come una cautela precauzionale volta a fin di bene. È vero che una prospettiva del genere è incompatibile con l'ideologia scientifica e democratica, ma ciò costituisce una confutazione/condanna di tale prospettiva solo per chi scambia la *fabula* della scienza e della democrazia con la realtà; cioè solo per chi, essendo preda della superstizione tecno-demo-cratice, attribuisce all'apparato oggi dominante un valore incondizionato, ed esclusivo.

In una situazione di questo genere, ragionare semplicemente da esseri umani “naturali” è veramente, secondo me, fortemente limitativo e preclude la possibilità di realizzarsi davvero come forme spirituali. Per dare una risposta piena alla nostra prima domanda dovremmo quindi avere il coraggio di alzare lo sguardo fino a prendere consapevolezza degli spiriti che guidano/governano davvero il sistema/apparato tecnologico. Cosa che richiede che il pensiero umano arrivi a scorgere, in ogni evento reale e storico, quella componente patica, affettiva, emotiva che sopra abbiamo evocato parlando del dolore, cioè la dimensione cosciente, relazionale, progettuale etc. che, come abbiamo visto, costituisce il fondamento veritativo dell’esistenza. Qualsiasi problema noi si voglia affrontare, anche quelli legati alla persona, non possiamo più fare a meno di porci questi problemi ultimi ma nello stesso tempo concreti.

8. A quali organismi spirituali apparteniamo e qual è la natura del corpo vivente?

È a questo punto che si presenta la seconda domanda posta sopra (nel § 6), cioè quella che chiede a quali organismi dello spirito apparteniamo noi di fatto. Ecco, credo che la questione che noi stiamo ponendo sia esattamente questa: quale è la “nostra” forma dello spirito? Anche per rispondere a questa seconda domanda è opportuno, oltre che necessario, rivolgere lo sguardo allo scenario mondiale. A prima vista la risposta è quella che scorge l’imporsi a livello globale di quella che ho chiamato la tecno-demo-crazia, ma il problema è che, venendo meno l’umano come un dato fisso e ineliminabile – e con ciò pure il kantiano umano come fine, e quindi anche i diritti umani come baluardo e faro della democrazia – viene meno pure la giustificazione etico-valoriale dell’agire tecnoscientifico basata appunto sull’intoccabilità della natura umana. E allora la tecno-demo-crazia si presenta a molti popoli della terra come una nuova religione particolare: la religione, appunto, del mondo occidentale. Intendiamoci, una grande, grandissima religione, paragonabile a quella degli Egizi; anzi, di gran lunga superiore a questa

dal punto di vista tecnologico. Il problema è quello, già evocato, che la nostra civiltà, nella misura in cui non si rende conto che il proprio fondamento di valore ultimo (l'uomo) è ormai venuto meno, si trasforma in una immane superstizione, come tutte le religioni che non sono consapevoli del loro essere basate su una fede. Per questo il vero pensiero è oggi quello che indaga le forme dello spirito che stanno dietro (o al fondo di) questa grande civiltà/(in-civiltà).

Tale pensiero, per avere successo, deve essere in grado di confrontarsi, oltre che con tutte le altre forme di sapienza e di spiritualità, anche con le concrete civiltà alle quali tali forme dello spirito donano vita. Senza peraltro che l'Occidente debba in qualche modo "rinnegare" la propria tradizione; semplicemente esso deve diventare consapevole di quali sono le forme effettive della soggettività/spiritualità che la sorreggono, e con ciò anche della sua relatività. A questo proposito mi torna in mente la battuta di Raimon Panikkar, sacerdote spagnolo ideatore della visione cosmoteandrica, il quale, avendo riscoperto, in Oriente, la propria radice indù, diceva: "Sono partito cristiano, mi sono scoperto Hindù e ritorno buddhista, senza cessare per questo di essere cristiano".

Rispetto agli enormi problemi che in tal modo si pongono mi limito qui a toccare due punti nodali, per fornire un'idea del compito che ci aspetta. Il primo riguarda la natura degli organismi ai quali danno vita le nuove forme dello spirito. È insomma la questione relativa a quelli che io chiamo "organismi spirituali"; termine con il quale mi riferisco alle forme dello spirito che hanno sì una realizzazione ("incarnazione") fisico-materiale, la quale però è libera rispetto alla fissità e ai vincoli imposti dalla natura propriamente corporea-animale.

Nell'epoca in cui lo stesso corpo dell'uomo diventa un problema si impone dunque la domanda circa la natura materiale-corporea delle nuove forme dello spirito, cioè appunto degli organismi spirituali. Si tratta di realtà fisico-materiali e "biologiche" le quali però, appunto in quanto guidate dallo spirito, hanno un principio di unità che è differente da quello fisico. Mentre il corpo dell'animale/uomo rimane sempre unito (nel senso che le sue membra restano sempre tutte insieme), l'organismo spirituale rimane unito anche quando le sue "membra" si

separano. La capacità dello spirito di distaccarsi dal corpo l'abbiamo già sperimentata ampiamente con l'uso prima dei libri, ora dei cellulari e così via. Ma nell'epoca in cui il corpo stesso viene sempre più generato dagli spiriti tecnologici esso assume una natura "fantastica", perché può comparire contemporaneamente in diverse parti della Terra e può incarnarsi in "corpi" differenti che però si possono anche riunire. Pensiamo (in maniera del tutto iniziale) ai gruppi umani che si organizzano legandosi in maniera anche molto forte ma senza il vincolo di una identificazione spazio-temporale. In questo senso l'espressione "tecnosauri", che altre volte mi è capitato di usare per descrivere certe forme di vita del tempo presente, è poco adeguata, appunto in quanto richiama troppo da vicino l'immagine del corpo animale, ed evoca con ciò piuttosto la figura del *Golem*. Qui forse l'espressione "corpo mistico" potrebbe essere più adeguata, ma, come sempre, va pensata libera dalla necessità che la vincola a una particolare religione o cultura, oltre che a una particolare forma corporea.

Ma in relazione al corpo c'è un altro aspetto decisivo che merita di essere sottolineato. In questo venir meno (*cedere*) di ogni necessità (*ne-cedo*), ciò che in particolare cade è il vincolo della morte e del dolore che pare definire il corpo animale. La possibilità di superare la "necessità" della morte e del dolore ci impone però di prendere consapevolezza del fatto che il corpo dell'uomo (e quindi l'uomo) è in quanto tale un "generatore di sofferenza". Noi siamo abituati a pensare che è quando le cose vanno male all'uomo che egli soffre; ma ora possiamo, e forse anche dobbiamo, vedere la cosa in una maniera diversa, alla quale accenno ora mediante un'immagine che riprende quanto scritto nel saggio precedente più volte citato. Come la lampadina genera luce grazie all'elettricità che in essa scorre, così l'uomo è un generatore di gioia grazie alla sofferenza che lo percorre, essendo appunto la sofferenza (*sub ferre*) la capacità di "sopportare" il dolore (infelicità) e con ciò di donare felicità alla vita. Per questo lo spirito accorto deve stare molto attento all'atteggiamento contemporaneo che, abbagliato dallo spirito faustiano della tecnica, si muove risolutamente nella direzione di eliminare il dolore (e la morte) dalla vita umana; perché questo rischia

di portare a una cancellazione della proprietà dell'umano, e del suo pregio specifico.

In gergo teologico potremmo allora dire: Dio crea l'uomo a sua immagine e somiglianza riconoscendosi quindi in lui; ma proprio in questo donare a un "altro" la sua stessa infinita libertà di creazione si realizza la possibilità del dolore. Ora, se noi togliamo all'uomo la possibilità del dolore (o, se vogliamo, "la necessità della possibilità" del dolore) gli togliamo anche la possibilità della gioia che consiste nella capacità di generare felicità mediante la "sopportazione" del dolore. L'uomo è pensato da Dio (uso queste metafore teologiche perché mi consentono di abbreviare molto il discorso, e non perché tutti si debbano sentire necessariamente vincolati ad esse) per dare luce al mondo, ma questa luce è generata dalla (possibilità della) sofferenza.

Come capite si tratta di temi di grande rilievo ma anche di grande difficoltà, ai quali comunque non si può più fare a meno di prestare attenzione proprio perché oggi l'uomo, con il suo corpo, non costituisce più un assioma indiscutibile e immodificabile.

9. A quale organismo spirituale vogliamo dare vita "noi"? La legge della vita come legge della libertà e insieme dell'amore, e la vera sapienza come scienza dei miracoli

Questa sintetica posizione del problema ci serve qui soprattutto per segnalare il dovere di individuare un esito positivo al pensiero che si interroga circa la natura degli organismi spirituali ai quali "noi vogliamo" dare vita perché li riteniamo belli, giusti e buoni (terza domanda posta nel § 6). A tal fine possiamo allora pensare a una forma di vita (un organismo spirituale, appunto), che scaturisca direttamente dalla legge della vita intesa come legge dell'amore, secondo quanto ci viene in qualche modo indicato da tutte le grandi tradizioni sapienziali del mondo.

La legge della "vita", infatti, può essere descritta come quella per la quale il vivente è caratterizzato dal fatto che il benessere o il malessere

(piacere o dispiacere) di ogni sua parte (ogni parte del suo organismo) è anche il benessere o il malessere di ogni sua "altra" parte. In tal modo le diverse, infinite forme che l'organismo spirituale assume sono definite dal fatto di essere una forma di vita nella quale tutti gli elementi che la compongono si riconoscono appieno in essa. Ciò accade però solo grazie allo spirito che resta libero rispetto a qualsiasi organismo vivente "dato/finito", e perciò anche rispetto al dolore che "definisce" il corpo del vivente in quanto animale-mortale. In tal modo la regola aurea della vita (fai agli altri quello che vorresti che gli altri facessero a te) giunge a compimento nella legge cosmica-umana-divina che dice: «Prenditi cura di ogni altro essere come di te stesso; perché – come ci insegna la sapienza indiana – *tat tvam asi* (tu sei quello), cioè *tu* sei la via, la verità e la vita». Affermazione che, peraltro, può naturalmente essere pensata in verità solo insieme alla consapevolezza del suo stare insieme all'altra verità, quella che dice: «*Io* sono la via, la verità e la vita».

Queste parole ci forniscono un'idea, sia pur sempre parziale e provvisoria, del compito che attende un pensiero che sia all'altezza della vita caratteristica della nostra epoca, compito che costituisce quindi il dovere pratico e operativo al quale è destinato lo spirito che anima la "nostra" vita attuale.

Nel viaggio lungo la via per la quale ogni sapere è "condizionato" e quindi revocabile, l'unica giustificazione etica è allora quella della "vera scienza", che integra la conoscenza scientifica e la connessa onnipotenza tecnologica con la "sapienza" che svela le "condizioni" alle quali anche l'operare della "*volontà di onnipotenza*" caratteristico della tecnologia (persino esso) garantisce il vero "bene-essere", cioè l'armonia e la soddisfazione dei viventi tutti. Il compito filosofico-sapientiale è appunto quello di determinare – anche concretamente – le "condizioni" che ne rendono possibile la realizzazione. Realizzazione che si presenta anche, ed essenzialmente, come un operare di tipo artistico, il quale però opera, oltre che con la natura e gli strumenti tecnici, anche con le forme di vita esistenti, a cominciare da quelle umane. In tale operare la sapienza giunge a compimento come "*scienza dei*

miracoli". Dove il primo miracolo consiste proprio nel comprendere in che senso una plateale contraddizione come questa – chi ha un atteggiamento scientifico non si aspetta miracoli, e viceversa chi confida in un miracolo non sta facendo scienza – possa esprimere qualcosa di assolutamente, perfettamente vero.

PER UNA INTELLIGENZA ARTIFICIALE RELAZIONALE.
ACCOMPAGNAMENTO ETICO DELL'EVOLUZIONE TECNOLOGICA

Leopoldo Sandonà

ABSTRACT

L'accompagnamento etico nel campo dell'IA può utilmente prendere spunto dai precedenti nella storia della bioetica. Attraverso il concetto di bioetica "traslazionale", che risponde alle derive di una medicina difensiva, è possibile anche per l'etica dell'intelligenza artificiale recuperare un triplice approccio di (a) ricerca e programmazione, (b) applicazioni, (c) formazione, in grado di verificare i presupposti e di integrare la prospettiva in essere.

Parole chiave: Accompagnamento etico, Medicina difensiva, Etica traslazionale, Vulnerabilità.

The ethical accompanying in the field of AI can be usefully oriented towards models taken from the history of bioethics. With the help of the concept of translational bioethics, reacting to trends in defensive medicine, the ethics of artificial intelligence can also find a tripartite approach consisting of research and programming (a), applications (b) and training (c), checking assumptions and continuously integrating the current perspective.

Keywords: Ethical accompanying, Defensive Medicine, Translational Ethics, Vulnerability.

«Questo è però il problema del destino dell'età della tecnica. In quest'epoca di illuminismo, nell'accrescersi del sapere e delle capacità, nella fede del progresso in sé, è divenuto spesso incomprendibile che cosa, per l'uomo, sia veramente importante»¹.

«Quando, nell'operare matematico, l'ignoto diventa l'incognita di un'equazione, è già bollato come arcinoto prima ancora che ne venga determinato il valore. [...] Ciò che appare un trionfo della razionalità soggettiva, la sottomissione di tutto ciò che è al formalismo logico, è pagato con la docile sottomissione della ragione a

¹ K. JASPERS, *Il medico nell'età della tecnica*, Raffaello Cortina, Milano 1991, 68.

ciò che è dato senz'altro. [...] Mentre il formalismo matematico, che ha per mezzo il numero, la forma più astratta dell'immediato, fissa il pensiero alla pura immediatezza. Si dà ragione a ciò che è di fatto, la conoscenza si limita alla sua ripetizione, il pensiero si riduce a tautologia. Quanto più l'apparato teorico si asservisce tutto ciò che è, e tanto più ciecamente si limita a riprodurlo. Così l'illuminismo ricade nella mitologia da cui non ha mai saputo liberarsi»².

1. Premesse

Anche a partire dalle provocazioni dei filosofi del diritto presenti in questo convegno è possibile indicare alcuni elementi di accompagnamento etico dell'intelligenza artificiale (d'ora in poi IA). Alcune brevi premesse sono doverose. Esse sono di duplice ordine. Da un lato questo intervento non intende essere esaustivo, perché siamo all'interno di definizioni *in progress*, di una *black box* i cui riferimenti sono spesso ignoti ai più, siamo dentro un itinerario aperto e quindi le stesse definizioni vengono continuamente aggiornate. Soprattutto a livello di esemplificazioni gli elementi risultano frammentati e spesso non decifrabili, sebbene non manchino anche nelle *Big Tech* percorsi in cui l'etica è centrale³. Dall'altro lato va sottolineato che il presente intervento si pone in continuità con il precedente convegno e con il contributo in esso offerto circa la possibilità di Comitati etici nell'IA. A partire dal parallelo tra mondo della bioetica classica medica e scenari dell'IA⁴ è

² M. HORKHEIMER-H. W. ADORNO, *Dialettica dell'illuminismo*, Einaudi, Torino 1997⁴, 32; 34.

³ <https://www.ibm.com/think/insights/ai-best-practices> (accesso 30 ottobre 2024). La stessa responsabile del comparto etico IBM, Francesca Rossi, ha offerto a più riprese una riflessione, anche etica, sul tema: F. ROSSI, *Il confine del futuro. Possiamo fidarci dell'intelligenza artificiale?*, Feltrinelli, Milano 2019; EAD., *Intelligenza artificiale: come funziona e dove ci porta la tecnologia che sta trasformando il mondo*, Laterza, Roma-Bari 2024.

⁴ Non mancano peraltro diversi altri paralleli tra bioetica classica ed etica dell'IA. In entrambi i casi si parte da documenti giuridico-politici, in un caso il *Belmont Report*, nell'altro l'*AI Act*. Inoltre nell'ambito dell'IA si richiamano spesso i classici quattro principi della bioetica – autonomia, beneficenza, non maleficenza, giustizia – cui viene aggiunto il principio dell'esplicabilità.

possibile allargare ulteriormente il ragionamento etico. Mentre però nel mondo seppur travagliato della bioetica si navigava in mari relativamente conosciuti, oggi navighiamo in oceani aperti o addirittura nello spazio. Le risposte non possono dunque che essere provvisorie e sottoposte a continua verifica.

2. Comitati etici, *ethical competence* e accompagnamento etico

Il punto di partenza della riflessione è la centralità dell'addestramento etico alle nuove tecnologie, in termini di *soft skills* spontanee che vanno verso una definizione di *expertise* morale e competenza etica che allarga il compito. Se ci si può fermare alle "abilità spontanee", tale strada presenta il problema di riferirsi a virtù che alcuni hanno e altri no. Si fa così un appello ad abilità che qualcuno possiede e altri non hanno. Le terminologie di *moral expertise* ed *ethical competence* sembrano allargare lo sguardo, andando verso una diffusione di quel pensiero etico-critico fondamentale per il futuro.

Da un lato ci troviamo di fronte all'*expertise* morale, cioè a persone che sappiano distinguere problemi nei vari contesti, esperti di etica che possono identificare e soppesare una vasta gamma di rischi etici, che hanno familiarità con concetti e distinzioni che sono preziosi nelle diverse scelte etiche; ci troviamo dunque di fronte a una competenza che riguarda la conoscenza delle teorie morali per risolvere i conflitti morali, le alternative dilemmatiche o meno.

In altri termini è l'*expertise* di un filosofo morale che abbia dalla sua anche esperienza di etica applicata e di familiarità con i luoghi etici della deliberazione. Attraverso l'*expertise* morale di questi professionisti possiamo arrivare a una competenza morale dell'organizzazione, non demandata ad alcuni ma come lievito che fermenta in tutta l'istituzione. Non si tratterà più solo di fare riferimento alle abilità innate che alcuni possiedono in ambito morale, ma anche queste diventeranno leva per una crescita organizzativa e sistemica. Come si può facilmente evincere, per arrivare a questo traguardo sarà fondamentale la formazione,

la creazione di spazi deputati per il confronto su tali temi e una verifica appropriata e doverosa. In questo trittico assume fondamentale importanza la presenza di chi detiene l'*expertise* morale in chiave individuale per divenire lievito di una *competence* etica condivisa⁵. Non ci basta una dimensione spontanea dell'etica ma abbiamo bisogno di esperti che ci permettano di arrivare a una competenza morale dell'organizzazione come lievito che fermenta in tutta l'istituzione.

3. Effetti del passaggio da "medicina difensiva" a "bioetica difensiva"

Nel continuare il confronto critico tra bioetica classica e mondo dell'IA è possibile cogliere un insegnamento relativo al tema della "medicina difensiva" e delle sue conseguenze nello strutturarsi di una vera e propria "bioetica difensiva", con il rischio che anche nell'IA prevalga un approccio procedurale, e quindi difensivistico, o al limite solo reputazionale. Il termine "medicina difensiva", riferito ad azioni motivate principalmente da questioni di contenzioso, ha una storia molto simile alla stessa definizione di bioetica, perché proviene dagli Stati Uniti e sul piano cronologico si è diffuso a partire dagli anni Sessanta. La medicina difensiva, intesa come l'insieme delle azioni cliniche che hanno lo scopo di proteggere il curante dai contenziosi o da qualsivoglia *outcome* avverso, può darsi come "positiva", laddove i medici prescrivano visite o esami superflui, e "negativa", quando i medici si rifiutano di curare pazienti con alto rischio clinico.

Dal punto di vista degli effetti consequenziali la medicina difensiva si caratterizza anzitutto attraverso la presenza di *test* di laboratorio non necessari e di *imaging* diagnostico o consulenze non necessarie, ma anche per l'utilizzo di procedure ad alto rischio sproporzionate in rapporto alla storia etico-clinica oppure, al contrario, per la riluttanza ad

⁵ Cf. L. SANDONÀ, "Luoghi etici" e competenze etiche nell'epoca dell'Intelligenza artificiale, in G. PIAIA-R. PRETE-L. STEFANUTTI (curr.), *Intelligenza Artificiale e tutela della persona umana. Implicazioni etico-giuridiche*, Triveneto Theology Press, Padova 2024, 130-131.

affrontare pazienti con alto rischio clinico. La bioetica difensiva che deriva dalla medicina difensiva, pur presentando delle differenze, ha paradossalmente fin da subito e in negativo, nel suo rapporto alla medicina difensiva, una componente organizzativa. Il risultato "a valle" pernicioso e dannoso della medicina difensiva in termini relazionali e organizzativi, e in ultima analisi allocativi ed economici, è una conseguenza che si rimpalla a livello organizzativo, dalla diagnostica alle prescrizioni e all'utilizzo della struttura ospedaliera stressata in modo eccessivo.

La bioetica difensiva, pur con differenze importanti, si dà come elemento della medicina difensiva. Se ci troviamo di fronte a un'implementazione della cultura bioetica nella bioetica tradizionale e nella bioetica traslazionale, la bioetica difensiva sulla scorta della medicina difensiva tende a prospettare un inaridimento della cultura bioetica, perché il risultato è un riferimento alla bioetica come ad un tassello accessorio e difensivo del sistema.

Nel concludere questo richiamo al passaggio dalla medicina difensiva alla bioetica difensiva, alcuni elementi emergono principalmente: nella problematica intravediamo anche una via di soluzione, poiché se la medicina difensiva e la bioetica difensiva toccano in negativo l'elemento organizzativo, che è un elemento di allargamento della bioetica medico-clinica tradizionale, in esso si cela anche una parte della soluzione.

4. Dalla deriva difensivistica alla "traslazionalità" in medicina e bioetica

Già a questo livello il mondo dell'IA può evitare di ricadere nel *loop* difensivistico mostrato dalla bioetica, forse non più in termini giuridici bensì in termini soprattutto tecnologici. Tuttavia proprio dalla reazione alla deriva difensivistica emergono interessanti indicazioni. In particolare si fa qui riferimento alla prospettiva della bioetica "traslazionale", che si radica nella prospettiva traslazionale della stessa medicina. Come si può opporre in questo senso la bioetica traslazionale ad una

dimensione difensivistica e di frammentazione del sapere bioetico?⁶ Se la medicina difensiva rappresentava il presupposto della bioetica difensiva, altrettanto possiamo affermare che la medicina traslazionale in chiave positiva sia la base della bioetica traslazionale. La medicina traslazionale è composta da un trittico di (a) laboratorialità-sperimentazione, (b) clinica-*bedside*, portando le scoperte al letto del paziente, (c) con una ricaduta anche di ambito comunitario e sociale. Aprendoci alla prospettiva di una bioetica traslazionale, questa si situa in una circolarità tra dimensione laboratoriale della medicina traslazionale, andando al *bedside* come consulenza etica e dimensione sociale. In chiave bioetica il laboratorio è il Comitato e la cultura etica organizzativa – oltre che le interazioni bioetiche con la prospettiva della sperimentazione –, il *bedside* è la consulenza etica così come ogni dimensione di appropriatezza clinica dell'intervento medico-sanitario, mentre l'aspetto sociale è la dimensione formativa e informativa, senza escludere le nuove forme di luoghi etici, come per esempio il cosiddetto "spazio etico". Per un verso la bioetica traslazionale si può incaricare di accompagnare la dimensione della medicina traslazionale, nel momento sperimentale (ad es. attraverso l'etica della sperimentazione), nel momento clinico (attraverso la consulenza etica e il comitato per la pratica clinica) e nel momento sociale, attraverso l'informazione e l'educazione. Ma c'è anche uno specifico etico della bioetica traslazionale, soprattutto a partire dalla crisi dei luoghi di deliberazione etica, attraverso i Comitati etici, la consulenza etica e l'apertura allo spazio etico come accompagnamento della medicina traslazionale.

Se la medicina difensiva rappresentava il presupposto della bioetica difensiva, possiamo al pari affermare che la medicina traslazionale in chiave positiva sia la base della bioetica traslazionale, con delle differenze e cercando di capire come la bioetica sul piano traslazionale

⁶ Per una rassegna di diverse prospettive di bioetica traslazionale cf. J.A. PARSONS-P. CAIRNS-J. IVES, *Translational bioethics*, in *Bioethics* 38/3 (2024) 173-176; E. LANPIER et al., *Special Issue on the Translational Work of Bioethics*, in *Perspectives in Biology and Medicine* 65/4 (2022) 515-709.

non possa essere semplicemente una fotocopia rispetto alla medicina traslazionale ma abbia delle peculiarità. La ricerca traslazionale cerca di applicare le scoperte della scienza di base per migliorare la salute e il benessere dell'essere umano. Molte fasi della ricerca traslazionale possono includere compiti non medici (tecnologia dell'informazione, ingegneria, nanotecnologia, biochimica, ricerca sugli animali, economia, sociologia, psicologia, politica e così via). L'applicazione dei comuni principi bioetici a queste aree potrebbe talvolta non essere fattibile, o addirittura impossibile.

Tuttavia, l'intero processo deve rispettare alcuni principi morali fondamentali. Il tritico di laboratorialità, portare le scoperte al letto del malato e avere un'apertura di ordine comunitario e socio-culturale richiama la prospettiva di una bioetica pragmatica, come risposta ad alcuni eccessi della storia della stessa bioetica. D'altro canto la bioetica traslazionale si offre come integrazione rispetto a una prospettiva della bioetica empirica, cercando di recuperare un equilibrio generativo con la bioetica tradizionale ed evitando di ridurre la stessa solo a un sapere scientifico-sperimentale. In questa direzione si correrebbe il rischio, dopo una bioetica solo filosofico-morale e una bioetica solo medico-legale, di rispondere a un eccesso con altri eccessi, fallendo la prospettiva integrativa.

L'aspetto laboratoriale dell'etica così intesa, tra prossimità alle persone assistite e apertura alla società, manifesta una circolarità generativa, perché dalla relazione con la persona assistita e dal rapporto con la società emergono poi nuove tematiche che torneranno al laboratorio e alla sperimentazione. Per la bioetica questo tritico si può intendere, nella vita del Comitato etico, come laboratorio sperimentale e prossimità alle situazioni di cura e terapia (che richiede non solo una proceduralizzazione ma anche flessibilità e una transizione competente di consulenza etica), fino al terzo elemento che è quello della società come spazio etico. L'idea è che ci sia un luogo etico sociale che possa condividere la dimensione bioetica classica e quella sociale, sia per accogliere delle istanze da parte della società, sia per fare di questo luogo un momento di *public engagement* in un'apertura di carattere organizzativo *ad intra*

ed etico-sociale *ad extra*⁷. Anche in questo caso, come per la bioetica difensiva, questo significa in chiave propositiva e integrativa dar corpo a tutta la dimensione della tematica del principio di giustizia; inoltre qui si rivela la dimensione etico-sociale della bioetica globale intesa come bioetica integrale, superando quindi anche da questo punto di vista una lettura limitante della bioetica globale in chiave solo geografica. Una ricerca traslazionale d'impatto richiede nuovi approcci alla bioetica, entrambi favoriti dall'adozione di strategie di coinvolgimento della comunità. Le conoscenze e l'esperienza della comunità affineranno la raccolta dei dati, la ricerca e le intuizioni accelereranno l'impatto delle applicazioni traslazionali dedotte per migliorare la salute individuale, il processo decisionale in campo medico e le politiche di sanità pubblica. Nel contesto della ricerca traslazionale con i *big data* sanitari, il coinvolgimento significativo della comunità dei ricercatori richiederà lo sviluppo e l'impiego di strumenti di *co-engagement* che a loro volta necessitano di una dimensione etica.

Dal punto di vista fondativo non va taciuta la rilevanza della prospettiva traslazionale⁸. Essa infatti costringe l'etica, nella sua traducibilità applicativa, a porsi il tema della propria identità profonda. Le strategie per colmare il divario tra teoria e pratica in bioetica devono essere a loro volta giustificate in termini etici, rappresentando lo specifico dell'impresa morale. In questo quadro la dimensione traslazionale incorpora la ricerca etica normativa, empirica e fondazionale, e continua a svilupparsi attraverso l'applicazione e di fronte a nuove sfide etiche. Il campo accademico della bioetica non ha forse ancora analizzato a sufficienza il proprio fondamento filosofico per capire come possa e debba essere rilevante dal punto di vista pratico, né abbia discusso in modo esauritivo i limiti degli impatti che i bioeticisti dovrebbero perseguire nelle organizzazioni. Di conseguenza, non c'è stata un'adeguata formazione

⁷ H. FERNANDEZ LYNCH-I.G. COHEN-B.E. BIERER, *Public Engagement, Notice-and-Comment Rulemaking, and the Common Rule*, in *IRB: Ethics & Human Research* 39/1 (2017) 17-19.

⁸ K. BAERØE, *Translational bioethics: Reflections on what it can be and how it should work*, in *Bioethics* 38/3 (2024) 187-195.

su come incidere in modo adeguato e appropriato sulle pratiche dei mondi professionali. Porre l'accento sulla traduzione del fondamento significa paradossalmente difenderne l'elemento "fondativo" e "speculativo". Che cosa fonda l'originario se non esprime estrinsecamente la propria dimensione? Se non ci sono obiettivi traslazionali, il fondativo finisce per inaridirsi, pur restando decisiva la dimensione intrinseca del sapere morale come riflessione epistemologica *ad intra*. La prospettiva traslazionale si pone come inveramento e insieme verifica della dimensione fondativa e speculativa.

5. La prospettiva traslazionale e l'IA

A partire dal confronto tra medicina difensiva e bioetica difensiva, ma anche tra medicina traslazionale e bioetica traslazionale, la ricerca di una riflessione etica sull'IA non parte anzitutto dalla valutazione generale della presenza di riferimenti etici in questo ambito, spesso lasciati alla formulazione importante ma necessariamente generica dei codici etici aziendali. Molte delle grandi multinazionali *tech* hanno principi molto simili alla bioetica, dall'autonomia alla giustizia fino alla beneficenza e alla non maleficenza. In particolare una direzione decisiva appare essere quella del principio di trasparenza nell'esplicabilità, poiché nelle forme complesse i sistemi devono essere leggibili⁹.

Come nel caso della medicina difensiva le pratiche mediche, legittime sul piano giuridico, possono manifestare una discrepanza molto forte dalla sostenibilità etica. Oggi chi opera in IA non opera "fuori" dalla legge ma "prima" delle leggi¹⁰. Divengono così fondamentali linee guida

⁹ L. FLORIDI, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina, Milano 2022, 91 ss., ove si parla esplicitamente di cinque principi dell'intelligenza artificiale in cui, ai quattro classici della bioetica, si aggiunge il quinto dell'esplicabilità; J. ADAMS, *Defending explicability as a principle for the ethics of artificial intelligence in medicine*, in *Medicine Healthcare and Philosophy* 26/4 (2023) 615-623.

¹⁰ L. PALAZZANI, *Dalla bio-etica alla tecno-etica: nuove sfide al diritto*, Giappichelli, Torino 2017, 57-58: «ci si aspetta molto dal diritto: forse più dal diritto che dall'etica, dopo le

per decisori e base per verificare in futuro questi fenomeni. Anche se in modo solo embrionale si può così cercare di capire come un accompagnamento etico può strutturarsi. L'etica dell'IA potrà così ritrovare in se stessa il trittico traslazionale già descritto, pur con le dovute differenze. Volutamente si indicheranno tre passaggi senza indicare una sorta di ordine fra i tre, ma ipotizzando diverse correlazioni. Il primo elemento indicato è quello della ricerca, della programmazione e dei principi, un momento per così dire interno; se coloro che determinano algoritmi e programmazione si sentono creativi¹¹, diversamente da come siamo abituati a pensare gli informatici, cioè in chiave computazionale e non "poetica", allora essi stessi hanno bisogno di un accompagnamento etico per far emergere i propri dubbi. Qui la problematica emergente è quella della libertà di azione del soggetto singolo o comunitario che programma: tale soggetto potrebbe essere vincolato da limiti aziendali nel far emergere i propri dubbi etici o potrebbe essere spinto da un approccio meramente operativo, non in grado di far emergere gli stessi dubbi etici. In questa fase, come ben mostrato dalla riflessione di Paolo Benanti¹², siamo nel cuore della cosiddetta algoretica, sia come riferimento ai principi che l'essere umano deve seguire per assicurarsi che i modelli di IA da esso sviluppati servano esclusivamente alla crescita del benessere sociale sostenibile, sia come tentativo di codificare all'interno del modello sviluppato quell'insieme di regole comportamentali che consentano alla macchina di assumere un comportamento rispettoso dell'essere umano e dell'ecosistema.

delusioni di lunghi dibattiti, rivelatisi a volte sterili e ripetitivi. Ci si aspetta la "certezza" della regolamentazione; si esige un punto di riferimento fermo per sapere finalmente come ci si debba comportare a livello collettivo (per sapere che cosa è "legale" e che cosa è "illegale"); e, avere certezze, contribuisce a rimuovere socialmente le paure collettive, a dissolvere i fantasmi e le angosce che a volte accompagnano la comunicazione (a volte, forse, troppo impulsiva), dei nuovi risultati raggiunti dalle conoscenze scientifiche e dalle applicazioni tecnologiche in biomedicina».

¹¹ Cf. intervento del prof. Paolo Moro.

¹² P. BENANTI, *Oracoli. Tra algoretica e algocrazia*, Sossella, Bologna 2018.

Far entrare l'etica dentro una *black box* altrimenti opaca significa compiere un nuovo passo dal mito alla ragione, senza la pretesa moderna di "illuminazioni" definitive.

Questo primo elemento a mio parere rappresenta la prima parte fondamentale dell'intervento etico, cioè della ricerca. Il secondo momento è quello applicativo in un'estensione molto ampia, nel mare aperto dei consumi e degli usi dell'IA. Come si può facilmente intuire, queste applicazioni legano il secondo momento a un'apertura immediatamente sociale. Rispetto alla medicina e alla bioetica traslazionale emerge quindi la prospettiva di un anticipo della prospettiva educativa e formativa propria del *public engagement* classico. Infatti l'utente si trova ad usare l'IA senza preparazione etica, mentre dovrebbe esserci da un lato un'ampia formazione ed educazione in materia, dall'altro una possibilità per l'utente di interagire e di poter manifestare i dubbi e i malfunzionamenti anche sul piano etico della tecnologia utilizzata. Inoltre nella fase applicativa la tecnologia mostra tutta la sua ambiguità, come indicato da Floridi, con una serie di minacce ma anche di opportunità che emergono in tale utilizzo¹³. Mentre in ambito medico e bioetico siamo di fronte ad applicazioni specifiche che in prima battuta riguardano delle specifiche professioni, l'ambito applicativo della traslationalità nell'IA ci rimanda quindi al mare aperto della bioetica globale, quasi richiamando le famose sei "P" di Potter (povertà, popolazione, inquinamento [*pollution*], progresso, pace, politica)¹⁴. In questo senso il concetto di bioetica traslazionale reversibile o circolare potrebbe essere interessante anche per l'IA, perché il rapporto tra applicazione e

¹³ FLORIDI, *Etica dell'intelligenza artificiale*, 182; 228. Per le minacce si vedano il commercio, i mercati finanziari, il furto di dati, il mondo delle droghe, i reati contro la persona, il furto e la frode compresa la sostituzione della persona. Sul piano delle opportunità si vedano l'ottimizzazione della tecnologia in campo ambientale, l'accompagnamento negli studi prevenendo abbandoni ed esiti negativi, l'assistenza per persone con disabilità cognitive, l'assistenza didattica individuale, il rilevamento di frodi.

¹⁴ V.R. POTTER, *Global Bioethics: Building on the Leopold Legacy*, Michigan State Univ. Press, East Lansing 1988, ripreso in H. TEN HAVE, *Bioetica globale. Un'introduzione*, Piccin, Padova 2020, 28.

ricerca si fa molto stretto e non può certamente essere unidirezionale ma circolare e almeno bidirezionale.

Il terzo elemento educativo e formativo si dà in questo senso come preliminare a entrambi ma anche come successivo: “preliminare” per preparare la popolazione come gli esperti a riconoscere le questioni etiche intrinseche al processo, ma anche “successivo” perché a partire dai processi si possono operare adeguate verifiche e prendere le adeguate contromisure. Spesso l’elemento formativo ed educativo è trattato come parte “residuale” del processo, ma sul piano di una disseminazione etica è in realtà il primo passaggio, in grado di accompagnare le applicazioni ma anche la programmazione.

Il modello traslazionale si offre come uno spunto in grado di interpretare un possibile accompagnamento etico. Senza dubbio per dar corpo a tale modello sarebbe fondamentale implementare le buone pratiche e le testimonianze, in grado di arricchire sul campo tale ipotesi metodologica. L’etica può porsi come accompagnamento se non la utilizziamo in chiave accessoria o eventuale, quando diritto o tecnica non possono più arrivare, o se non la utilizziamo come riflessione personale in chiave meramente individuale. Non abbiamo quindi bisogno solo di Comitati etici ma anche di una dimensione più vasta e traslazionale che impegni ricercatori, programmatori, decisori e utenti/cittadini in una piena apertura sociale, considerandoci tutti come soggetti vulnerabili nel contesto iper-tecnologico, ma dando a questa vulnerabilità anche un significato proattivo e propositivo. Come afferma Ten Have,

«piuttosto che avere Comitati etici che rappresentano soggetti vulnerabili e che parlano per loro, questi argomenti devono essere inclusi in modo che essi possano parlare da soli. I Comitati etici non saranno quindi semplicemente meccanismi procedurali, ma piattaforme per l’*empowerment* dei gruppi vulnerabili»¹⁵.

¹⁵ Id., *Vulnerabilità. Una sfida per la bioetica*, Mimesis, Milano-Udine 2024, 382.

6. Conclusioni aperte

Tornando in conclusione alle citazioni iniziali, dopo aver offerto qualche spunto integrativo secondo il modello della bioetica traslazionale per integrare la mera prospettiva dei Comitati etici come strumento di accompagnamento etico, possiamo affermare che, perché il potere nelle nostre mani non diventi fuori controllo e quindi preda di poteri dominanti, è fondamentale strutturare un adeguato accompagnamento etico che comprenda anche elementi antropologici e contestuali. In questa chiave la lettura guardiniana di un' "ascesi del potere", in cui quest'ultimo sia compreso anche in chiave dinamica e verbale, appare un punto interrogativo pregno di futuro e di possibilità interessanti, e non gravido di apocalittiche, distruttive minacce. Con questo punto interrogativo lasciamo le indicazioni definite nel contributo come aperte e generative per ulteriori possibili approfondimenti:

«L'uomo è l'essere a cui Dio ha dato il potere sul mondo e su se stesso e da cui ha imposto la corrispondente responsabilità? [...] Quando il potere diventa demoniaco? Lo è forse già oggi, se si considera la sua sfrenata crescita nella scienza, nella tecnica, nella possibilità di influsso sociale ecc., nonché la minacciosa spensieratezza [*Sorglosigkeit*] che predomina in questa situazione? [...] C'è un'ascesi del potere?»¹⁶.

¹⁶ R. GUARDINI, *Domande sul problema del potere*, in ID., *Scritti politici (Opera omnia, VI)* Morcelliana, Brescia 2005, 599.

NOI E L'INTELLIGENZA ARTIFICIALE. DAL PATHOS DELLA PAURA ALLA CONFIDENZA CRITICA

Francesco Totaro

ABSTRACT

Verso la IA è diffuso un sentimento di paura per la minaccia di cancellazione dell'umano, al quale soltanto si ritiene si possa attribuire intelligenza. In realtà si può parlare anche di una intelligenza computazionale. La differenza tra l'umano e la macchina è dovuta al carattere di apertura all'"intero" della coscienza umana. Pertanto, occorre subordinare i fini incorporati negli strumenti artificiali e nei loro processi all'ordine delle finalità umane. Su questa via è possibile passare dalla paura alla confidenza critica grazie anche a maggiori competenze conoscitive, in un rapporto dialettico con l'artificiale che deve sempre misurarsi con i rischi di condizionamento.

Parole-chiave: intelligenza artificiale, paura, coscienza dell'intero, confidenza critica.

Towards AI there is a feeling of fear for the threat of deletion of the human, to which only one believes that intelligence can be attributed. In fact, we can also talk about a computational intelligence. The difference between human and machine is due to the character of openness to the whole of human consciousness. Therefore, the purposes embodied in artificial instruments and their processes should be subordinated to the order of human purposes. On this path it is possible to pass from fear to critical confidence thanks also to greater cognitive skills, in a dialectical relationship with the artificial that must always measure up to the risks of conditioning.

Keywords: artificial intelligence, fear, consciousness of the whole, critical confidence.

1. L'intelligenza artificiale e il vissuto umano

Il vissuto umano è fatto di azioni e passioni. Azioni e passioni sono strettamente intrecciate. Le passioni possono spingere ad agire e l'agire

ha spesso la sua radice in una passione. Per quanto azioni e passioni siano non di rado tra loro indiscernibili, alle passioni si attribuiscono i caratteri della spontaneità e immediatezza, anche quando hanno un'estensione duratura nel tempo. Le passioni danno luogo a sentimenti ed emozioni che permeano la sensibilità diffusa. Se chiamiamo tutto ciò *pathos* presente nella dimensione sia individuale sia collettiva, possiamo chiederci quale sia la sua tonalità prevalente nei confronti della evoluzione estrema dell'artificiale che abbina ad esso l'intelligenza, considerata prerogativa esclusiva dell'umano. Per difendere una tale prerogativa, non pochi si sentono in dovere di censurare l'associazione del crisma dell'intelligenza a un artificio che si limita a operazioni di calcolo. Certamente – come più avanti vedremo – non mancano i motivi di distinzione, se però "calcolare" è parte dell'*intelligere*, e se il calcolo può essere intriso di input in cui vengono tradotti e travasati contenuti pertinenti all'umano grazie all'applicazione di una logica combinatoria di tipo binario (discontinua e di accumulazione statistica in tempi limitati), allora la pertinacia con cui si rifiuta ogni capacità di intelligenza alle macchine e ai robot si presterebbe a essere interpretata come pregiudizio difensivo. Un tale pregiudizio – ci chiediamo – è forse alla base di una sorta di *iper-pathos* o *super-pathos* che sovrasta i molteplici *pathos* o la gamma delle *affezioni* particolari riguardo alla Intelligenza artificiale?

2. Il *super-pathos* della paura

Pathos è un lemma polivoco: nella semantica del *pathos* che si connette alla lingua dei greci si va dall'accezione più neutra di *fenomeno* (manifestazione di qualcosa che accade) ad altre con connotazione negativa (sciagura, disgrazia, terremoto) e, relativamente all'umano, si mette in evidenza una scala di affezioni e sentimenti fino alla passione alla quale si cade in preda (per es. la folle passione amorosa). Sappiamo bene peraltro che, tra i filosofi antichi, al *pathos* è stata indicata la via della "purificazione" (*catarsi*) e, come rimedio radicale al suo dominio, l'etica della liberazione dalle passioni (l'*apatheia* e l'*atarassia*).

Riprendendo allora la nostra domanda, si dà un *super-pathos* verso l'intelligenza artificiale? E in cosa consiste? Nella "paura" della esteriorizzazione o alienazione dell'umano, che spinge alcuni fino alla rimozione del fenomeno, al punto da dire che l'intelligenza artificiale non esiste (non è né sostanza né accidente, avrebbe detto il noto don Ferrante, oppure in versione mitigata: è forse un accidente, ma non è affatto sostanza). La paura dell'artificio si esprime nella domanda assillante: cosa – di irrefrenabile e irresistibile – l'artificio può fare di noi? Su questa via matura e si consolida appunto il *super-pathos* dell'artificio come minaccia assoluta.

Non è che non ci siano motivi per la configurazione dello stereotipo dell'artificio equiparato a una minaccia assoluta. Potremmo persino riconoscere che il *timor artificii* è un *initium sapientiae* e che l'etica della paura, per fare il verso a Hans Jonas, può ispirare l'etica della responsabilità¹. Volendo mettere le cose in positivo, potremmo dire che il timore dell'artificio ci può allertare o darci la scossa. Dandoci le energie per fare del *pathos* "timoroso", rispetto al nuovo che ci spiazza, un punto di partenza che ne rovescia il senso negativo di "sciagura" e lo riconfigura come "fenomeno" o accadimento antropologico-ontologico. Un fenomeno che esige quindi una "qualificazione", cioè un orientamento qualitativo conforme alla intenzionalità umana.

3. La domanda giusta e il carattere irriducibile della intenzionalità umana

La domanda va allora rovesciata: non "cosa l'artificio può fare di noi", bensì "cosa noi, noi umani, possiamo fare dell'artificio"? Qui siamo sfidati a un compito duplice: di ricomprensione teorica e di orientamento pratico. Qui si impongono pure esigenze di approfondimento

¹ A questo *alert* possono essere utili analisi che si collocano sul versante della "paura", come quella del notissimo Yuval Noah Harari sulle colonne del *Financial Times* del 26 ottobre 2024, dal titolo *Beware the AI bureaucrats*, dove si denuncia specialmente il trasferimento di poteri dagli umani alle piattaforme algoritmiche dei *social media*.

e di fondazione o ri-fondazione. Anzitutto occorre ricalibrare lo statuto dell'intenzionalità umana rispetto alla funzione della macchina e chiedersi se (e perché) l'intenzionalità umana non sia riducibile a una sequenza di operazioni algoritmiche, quale che sia il loro livello di incorporazione dei dati e di elaborazione degli stessi nel calcolo informatico (*computing*) per la soluzione di problemi. Andando al fondo della questione, si tratta di affermare il carattere irriducibile della intenzionalità umana ai processi di calcolo dell'artificio. Questo permette di argomentare l'infondatezza di ogni tipo di riduzionismo connesso alla equiparazione della coscienza (aperta all'"intero" in modo formalmente illimitato) al funzionamento neuro-cerebrale che le fa da supporto.

La pretesa riduzione dell'intenzionalità umana al funzionamento della macchina è infatti un capitolo particolare del riduzionismo neuro-cerebrale, che fa della coscienza l'effetto di funzioni fisiologiche in un rapporto biunivoco, allorché le funzioni neuro-cerebrali non sono considerate "concomitanti" con l'attività coscienziale bensì causalmente "determinanti". L'intenzionalità "interale" della coscienza umana – coscienza aperta intenzionalmente al tutto, senza limiti – viene così assorbita, in queste impostazioni, nel funzionamento "macchinico" degli organi neuro-cerebrali. Beninteso, non si vuole indulgere a una vaga e superficiale reazione di insostenibile stampo antiscientifico; occorre piuttosto intessere un'alleanza con gli studi scientifici che si discostano decisamente da impostazioni riduzionistiche. Mi permetto di fare qualche incursione in proposito. Trovo interessanti le tesi di Marcello Massimini e Giulio Tononi, i quali affermano che in ogni caso il cervello ha davvero qualcosa di speciale e i suoi ingredienti fondamentali sono l'informazione e l'integrazione: «Un sistema fisico è cosciente nella misura in cui è in grado di integrare informazione»; pertanto l'informazione «non vale nulla se non si accompagna all'altro ingrediente fondamentale: l'integrazione»².

Questi autori non mancano di sottolineare le affinità con il filosofo della coscienza David Chalmers. Per Massimini e Tononi la coscienza

² *Ibid.*, 1105.

corrisponde quindi alla massima concentrazione di informazione, alla «massima informazione che un sistema genera al di sopra delle parti che lo compongono, siano queste atomi, molecole, neuroni, singoli circuiti o aree corticali. Noi siamo questo nocciolo duro di informazione irriducibile e da questo dipendono le nostre scelte – da niente di più e da niente di meno»³. La coscienza, possiamo dire, è eccedenza di informazione integrata che permette una causazione libera, che non si riduce a subire una semplice successione indipendente dalla nostra decisione cosciente. E una tale eccedenza ha luogo nel sistema talamocorticale integrato di un “singolo” individuo, l'unico luogo dove le relazioni esterne diventano inscindibili relazioni interne. Ritengo che questa tesi di irriducibilità possa essere riferita anche al rapporto tra coscienza umana e intelligenza artificiale.

Conosciamo pure la posizione di Federico Faggin. Sebbene sia discutibile la sua estrapolazione “metafisica” della fisica quantistica in versione spiritualistica, è fecondo il paradigma coscienziale che egli suggerisce. Faggin distingue la «elaborazione cosciente del significato», basata principalmente su operazioni associative e analogiche tra i *qualia* insieme con i processi logici e lineari della mente razionale, dalla «elaborazione automatica dei simboli» a mezzo di macchine, come i dispositivi dell'Intelligenza artificiale⁴. È grazie alla «supervisione della coscienza» che noi connettiamo l'informazione con i significati che sono alla base della nostra esperienza vissuta, conservata nella «nostra memoria a lungo termine». Non possiamo insomma appiattare la memoria umana sulla memoria del computer: non si può «eliminare l'abisso che separa gli esseri umani dalle macchine cosiddette intelligenti»⁵.

Queste posizioni non riduzionistiche sono “compatibili” con la visione della coscienza “trascendentale”, intesa come apertura a un “intero” irriducibile ai contenuti che si succedono nelle situazioni sempre

³ *Ibid.*, 2787.

⁴ Cf. F. FAGGIN, *Irriducibile. La coscienza, la vita, i computer e la nostra natura*, Mondadori, Milano 2022, 93-94.

⁵ *Ibid.*, 92.

parziali alle quali la coscienza si applica. E allora, riprendendo il filo della riflessione, a voler cogliere la nota diversificante dell'esperire e del vivere umano rispetto ad altri modi di sperimentare e di vivere, la individuerei appunto nella "coscienza dell'intero", cioè nella coscienza aperta intenzionalmente al tutto senza limiti. Si tratta di un'apertura che riguarda anzitutto la forma dell'intenzionare e non presume il pieno possesso dei contenuti ai quali questo intenzionare si dispone.

4. L'"artificio" nell'ordine delle finalità umane

Il pensiero e l'agire degli umani non sono però autoreferenziali e sono da sempre intrecciati con la dimensione dell'"artificio". Gli artifici, al plurale, vengono pensati e "agiti" dagli umani. La nostra preoccupazione, non da oggi ma oggi cresciuta in modo esponenziale, è che, oltre una certa soglia, possano essere gli artifici ad "agire" gli umani (è la storia dell'apprendista stregone che non riesce più a controllare gli spiriti da lui stesso evocati). Nella situazione attuale la preoccupazione è, infatti, che gli artifici si sostituiscano agli umani addirittura nell'attività del pensiero cosciente e autocosciente. A causa di una tale preoccupazione, persino chiamare "intelligente" una macchina sembrerebbe avallare la confisca della qualità suprema dell'umano, il pensiero, da parte dell'artificio. Con l'aggravante che una tale invasione di campo minaccerebbe di coinvolgere pure la dimensione corporea complessiva degli umani, alterandola o persino eliminandola. Qui si agita lo spauracchio del post-umano e del trans-umano.

E allora: quale tutela per l'umano in preda alla paura dell'artificiale? Un approccio non pauroso al sistema degli artifici esige anzitutto un esercizio superiore di pensiero cognitivo e, al tempo stesso, una disponibilità non generica a scomporre il sistema complessivo delle macchine nelle molteplici articolazioni che lo costituiscono. La "nostra" intelligenza delle macchine ci doterebbe della capacità di controllare le macchine non solo nei loro esiti a valle, ma anche nella loro impostazione a monte (e qui sarebbe più che opportuno l'incontro tra filosofi e tecnici).

Occorre fare un passo ulteriore per la ri-comprensione dell'artificiale. Le prospettive offerte dall'avanzamento della tecnica possono essere incardinate nella prospettiva più ampia della "persona umana" come prospettiva delle prospettive, riaffermando quindi nell'umano la capacità prospettica "interale" e valorizzando il sistema delle macchine come prospettiva parziale. Si tratta in altri termini di portare l'ordine delle finalità incorporate nel sistema delle macchine (nel loro funzionamento le macchine ormai perseguono fini e non sono più semplici mezzi, come un filosofo italiano della cosiddetta "documanità" si ostina a ritenere) a disporsi come ordine "infravalente" rispetto all'ordine delle finalità dell'umano come intero. A questa sfida e a questo compito non possiamo sfuggire.

5. L'intreccio dialettico tra umano e artificio e i suoi rischi

L'intreccio tra "ciò che noi possiamo fare dell'artificio" e "ciò che l'artificio può fare di noi" è decisamente dialettico, quindi anche conflittuale. L'artificio può fare molto di noi: può essere invasivo e pervasivo, producendo dipendenza al punto tale da esercitare la costrizione, anche invisibile e inavvertita, ad adeguare noi stessi ai modelli elaborati dall'artificio. E qui bisogna fare attenzione a una sorta di gioco degli specchi, evidente nell'ambito delle invenzioni e delle applicazioni della Intelligenza artificiale: inizialmente può essere la macchina a riflettere l'immagine dell'umano cioè a conformarsi all'umano per replicarne comportamenti e prestazioni, ma strada facendo l'immagine dell'umano può diventare, o "deve" diventare, proprio quella che la macchina a lui riflette di rimando. L'umano si restringerebbe così nel profilo "calcolante" che gli viene restituito dalla macchina. Si darebbe il passaggio dal robot umanizzato all'umano robotizzato. Gli umani potrebbero permettersi di agire e di operare solo nella misura in cui la macchina lo permette, in base a codici diventati non modificabili e indisponibili per gli umani in carne ed ossa.

Su questa via, la macchina da esecutiva diventa "normativa". E la normatività della macchina in che rapporto starebbe allora con l'autonomia dell'umano e il suo campo di scelta? Tutto ciò che non si adegua alla normatività della macchina o, più banalmente, non è "accettato" dalle prestazioni della macchina, diventerebbe irrealizzabile e persino impossibile? Ci può servire il riferimento a un autore competente e acuto come Luciano Floridi, il quale suggerisce a ragione di non sopravvalutare il robot visto che, tra l'altro, non riesce ad allacciare le scarpe, operazione che riesce anche a un bambino. Ma l'esito probabile di questa insufficienza – a nostro avviso – può essere la eliminazione delle scarpe con i lacci, per conformare gli *skills* umani alle funzioni di cui il robot è capace... È solo un esempio di cancellazione delle prestazioni umane cui la macchina può indurre. Chiediamoci allora: di fronte al rischio di emarginazione e di spiazzamento antropologico dell'artificio, come tenere fermo il timone e far prevalere il nostro "poter fare" nei confronti dell'artificio? Forse non c'è una risposta uniforme e sul punto occorrerebbero analisi anche di tipo empirico.

6. Dalla paura alla confidenza critica

Tiriamo i remi in barca. Ho portato l'attenzione sulla questione di fondo: l'irriducibilità della coscienza al funzionamento della macchina. Mantenere ferma questa irriducibilità è la condizione di possibilità fondamentale per essere noi a governare l'artificio e per farlo anche con la dovuta umiltà, consapevoli di non essere onnipotenti ma di avere sempre e costantemente bisogno dell'artificio per essere umani e diventare più umani. Allora, al *super-pathos* della paura potrebbe subentrare il *pathos* di una ragionata "confidenza", accompagnata dialetticamente da una coscienza "critica". Qui entra in campo un'etica del limite e della misura, capace di accettare un decentramento antropologico da parte dell'umano, peraltro sempre in atto nella sua storia, e di sottrarre però l'umano al determinismo, sia diretto sia di riflesso, dell'artificio. Se infatti l'artificio è sempre "condizione" imprescindibile dello sviluppo

dell'umano, esso è, al tempo stesso, "condizionante" nei suoi confronti. In sintesi, occorre imboccare un cammino virtuoso che superi la pretesa del dominio antropocentrico, ma non cada nella trappola di un decremento ontologico indotto dalla smisuratezza tecnologica. Penso che questo sia un traguardo proponibile e auspicabile di "saggezza".

VISUAL CULTURE E LOGOFOBIA NELL'ERA DIGITALE

Flavia Silli

ABSTRACT

L'IA sta trasformando lo *status* antropologico, privilegiando un approccio "visuale" alla conoscenza. L'uso delle tecnologie digitali modifica l'esperienza relazionale e l'interazione con la realtà fisica attraverso un *medium* rappresentazionistico. Per affrontare la pervasività tecnologica risultano preziosi sia il riconoscimento della intenzionalità dell'immagine, sia la promozione educativa della "logofilia" personalista di M. Sciacca e di L. Stefanini, che riconduce ogni espressione della produttività umana alla matrice onto-assiologica della persona.

Parole-chiave: Estetizzazione, autoreferenzialità, "logofobia", "iconofilia", estetica personalista.

The AI is transforming the anthropological status, favouring a visual approach to knowledge. The use of digital technologies modifies the relational experience and interaction with physical reality through a representational medium. In order to deal with the pervasiveness of technology, both the recognition of the intentionality of the image and the educational promotion of the personalist logophilia of M. Sciacca and L. Stefanini are valuable. Their perspectives allow every expression of human productivity to be traced back to the onto-axial matrix of the person.

Keywords: Aesthetisation, self-referentiality, logophobia, iconophilia, personalist aesthetics.

Questo contributo prende le mosse da una riflessione del sociologo Jean Baudrillard che diversi anni prima della "rivoluzione digitale" scrisse:

«L'età della simulazione comincia con l'eliminazione di tutti i referenti, peggio: con la loro resurrezione artificiale in un sistema di segni, che sono una materia più duttile dei significati perché si prestano a qualsiasi

sistema di equivalenza, a ogni opposizione binaria e a qualsiasi algebra combinatoria [...] È piuttosto una questione di sostituzione del reale con segni del reale; [...] sarà un iperreale al riparo da ogni distinzione tra reale e immaginario»¹.

L'*input* più significativo di tale osservazione riguarda a mio avviso la distinzione epistemica fra "sistema di segni" e "significato", che ci aiuta a comprendere un fenomeno presente oggi nella *visual culture*²: l'iper-realtà come potenziamento delle virtualità umane dove si dissolve il limite etero-referenziale³. È significativa a questo riguardo l'accezione semantica attribuita da Baudrillard alla nozione di *simulacrum*, che viene denotato come "ciò che ha sostituito la realtà". Mi propongo qui di mostrare come l'estetica personalista, risalendo ai fondamenti epistemici del rapporto tra "immagine", "realtà" e "soggetto", offra preziosi spunti sia per tracciare la via di una loro costitutiva connessione diversificata, sia per contrastare efficacemente l'estetizzazione antropologica che si esprime nel primato dell'apparire sull'essere.

Per tematizzare la corretta interazione fra l'essere *online* e l'essere *offline*⁴, tra il virtuale e il fisico, occorre enucleare e tentare di sciogliere un nodo critico particolarmente significativo: l'assolutizzazione della rappresentazione in termini puramente narrativi ed autoreferenziali, che impedisce di coglierne l'*intentio* onto-semantica. Da questo punto di vista risulta molto interessante l'interpretazione critica della tecnologia immersiva argomentata da Baudrillard, che individua nella simulazione virtuale non tanto una de-realizzazione ma piuttosto un eccesso di realtà e di informazioni che rendono passivo il soggetto e

¹ Citato in AA.Vv., *Il libro della sociologia*, Gribaudo, Milano 2018, 199.

² Questa espressione trae ispirazione dal libro di N. MIRZOEFF, *Introduzione alla cultura visuale*, a cura di A. Camaiti Hostert, Meltemi, Roma 2021.

³ Cfr. l'interessante saggio di D. BONDI, *Come volontà e rappresentazione, appunti etico-antropologici sulla realtà virtuale*, in AA. Vv., *L'umano e le sfide della tecnica*, Morcelliana, Brescia 2019, 119-126.

⁴ Ho richiamato di proposito l'efficace espressione coniata da L. Floridi per evidenziare gli effetti epistemici della rivoluzione digitale. Cf. L. FLORIDI, *Pensare l'infosfera*, Raffaello Cortina, Milano 2020.

che provocano una sorta di “stènosi” alla sua capacità critica. Mentre si può accogliere come legittima sul piano epistemico l’attribuzione del termine “esperienza” alla dimensione virtuale, più controversa risulta a mio avviso la denotazione di “reale” all’immagine digitale. Come sottolineato opportunamente da Roberto Diodato, se lo spazio e il tempo delle tecnologie inter-attive multimediali risultano condizionate dal totale coinvolgimento e dall’identificazione dell’utente con il *medium*, la questione dell’identità, anche se «è al tempo stesso decorporeizzata e ipersensibilizzata»⁵, riemerge con forza sul fronte assiologico ed “aletico”, dal punto di vista cioè del valore e dell’attendibilità della rappresentazione. È chiaro che in un modello di estetica costruttivista questo problema non si pone, in quanto «la realtà virtuale [...] è connotata in termini [...] di una *creatio mundi*»⁶ dove tutto si può realizzare di fatto, anche cose che sono concepite in opposizione l’una all’altra, *in primis* il mondo reale da una parte e l’immaginario dall’altra. Nella dimensione virtuale il visore viene assorbito in una unità indistinta con lo schermo, in una sorta di monismo dissolutore di ogni riferimento estrinseco o etero-centrato. Come afferma con rigore analitico e argomentativo Baudrillard,

«Lo schermo è un luogo di immersione, ed ovviamente di interattività, poiché al suo interno si può fare quel che si vuole; ma in esso ci si immerge, non si ha più la distanza dello sguardo, della contraddizione che è propria della realtà. In fondo tutto ciò che esisteva nel reale si situava all’interno di un universo differenziato, mentre quello virtuale è un universo integrato. Di certo qui le care vecchie contraddizioni fra realtà e immaginazione, vero e falso, e via dicendo, vengono in certo modo sublimite dentro uno spazio di iper-realtà che ingloba tutto, ivi compreso un qualcosa che sembrava essenziale come il rapporto fra soggetto e oggetto»⁷.

⁵ R. DIODATO, *Estetica del virtuale*, Mondadori, Milano 2005, 20.

⁶ T. MALDONADO, *Reale e virtuale*, Feltrinelli, Milano 1992, 78.

⁷ Cfr. <https://www.repubblica.it/online/internet/mediamente/ baudrillard/ baudrillard.html>.

Per comprendere l'atteggiamento improntato all'estetismo che qualifica la *visual culture* ho tratto beneficio dalla lettura di una suggestiva e lungimirante riflessione di Michele Federico Sciacca, contenuta nel libro *Gli arieti contro la verticale*. In riferimento al consolidarsi del primato della cultura dell'immagine, egli scrive:

«L'infantilismo cronico [...] è ciò a cui mira la civiltà della tecnica e del benessere, l'anticultura. Il pensiero è mediazione, la parola è logos: l'immagine immediata dispensa dal travaglio della riflessione e della meditazione, dallo sforzo della parola propria, che è sempre creativa perché rivelativa dell'essere, mentre non lo è il termine esatto che è il linguaggio della scienza. Una pura cultura dell'immagine è davvero la morte del pensiero, che forse è il sogno del sonno in cui è caduto l'uomo odierno. [...] andiamo verso un'epoca oscura [...] un'epoca che proverà fino in fondo tutta la miseria incivilita dell'uomo senza verità e senza Dio, estrema indigenza spirituale»⁸.

Sciacca diagnosticò precocemente uno dei disagi della tecnocrazia occidentale: la *logo-fobia*, improntata all'indifferenza per la *ratio* sapienziale e al disimpegno etico e atletico. La perdita del *lògos*, come acutamente prefigurato da Sciacca, genera quell'anticultura che è oscuramento dell'intelligenza e nella quale l'essere umano, armato della tecnica più sviluppata e pur disponendo di tutti i mezzi per vivere, dispera di ragioni ultime e definitive per esistere. Secondo questa prospettiva il disimpegno nella ricerca di senso dell'immagine (anche quella virtuale-immersiva) determina un'alterazione del rapporto triadico soggetto-immagine-realtà.

Nel processo di estetizzazione contemporanea della realtà che ho qui messo a fuoco si può intravedere uno sradicamento reificante dell'immagine dall'atto generativo della persona. La sfida da accogliere è la ricerca di una corretta ermeneutica del nesso realtà-immagine per sventare efficacemente sia l'insidia "idolatrice" (che assolutizza l'immagine in se stessa), sia quella "icono-clastica" (che rifiuta l'attendibilità comunicativa dell'immagine). Questa terza via *iconofila* è quella tracciata e intrapresa dall'estetica personalista (egregiamente rappresentata da

⁸ M. F. SCIACCA, *Gli arieti contro la verticale*, Marzorati, Milano 1969,148-150.

Luigi Stefanini) secondo la quale l'immagine non definisce l'essere, ma lo "in-finisce", esprimendolo incompiutamente.

Diversamente da Baudrillard, che tende a sottolineare l'inevitabile "degenerazione" dell'*eikòn* in *eìdolon* e a snaturare il ruolo epistemico dell'immagine, qualificandola come alterazione "simulacrica" che dissimula l'oggetto rappresentato, Stefanini assegna all'immagine la funzione di cerniera e di *medium* tra il significante e il significato, capace di coordinare e di trasmettere, in una sintesi simbolico-concettuale non esaustiva, le molteplici declinazioni prospettiche dell'oggetto rappresentato. Se rapportiamo questo discorso al fenomeno dell'identità in rete o della costruzione della *self identity* nei social media, ci rendiamo conto della rilevanza che assume l'*iconofilia* imaginistica di Stefanini. Da un lato infatti, la funzione significante dell'immagine implica sia una considerazione del legame con il soggetto che la produce, sia un riferimento alla dimensione di ulteriorità e alla ricerca di completezza della realtà che essa rappresenta solo in parte; dall'altro risulta impossibile sul piano epistemico costruire la propria identità attraverso una narrazione visiva⁹, perché l'immagine prodotta dal soggetto non contiene mai tutto l'essere che, per usare un'espressione cara a Gabriel Marcel, resta e resterà sempre "mistero"¹⁰. Questa sproporzione – argomentata con rigore epistemico da Stefanini nel suo *Imaginismo come problema filosofico*¹¹ – tra la funzione semantica dell'immagine e l'essere cui essa rinvia, smaschera le criticità insite nel processo di mediatizzazione sradicato dalla referenzialità ontologica e inter-soggettiva. Tutto il "dietro" che permane al fuori del campo rappresentativo dell'immagine deve sospingere il destinatario della comunicazione visiva ad ascendere verso un ulteriore sforzo di comprensione. Per spiegare ai nativi digitali che le immagini del proprio profilo *social* non definiscono l'identità, occorre decostruire il carattere auto-referenziale dell'immagine

⁹ S. TOSONI, *Identità virtuali, comunicazione mediata da computer e processi di costruzione dell'identità personale*, Franco Angeli, Milano 2004.

¹⁰ Cfr. G. MARCEL, *Il mistero dell'essere*, a cura di G. Bissaca, Borla, Roma 1987.

¹¹ Cfr. L. STEFANINI, *Imaginismo come problema filosofico*, Cedam, Padova 1936.

come costruzione di senso, determinato dal ripiegamento del rapporto soggetto-immagine su sé stesso che snatura quella dialettica costitutiva, capace di rinviare ad un'ulteriorità estrinseca rispetto all'oggetto significato.

A conclusione di questa breve riflessione risalta il valore irrinunciabile di una formazione integrale e permanente della persona, *condicio sine qua non* della rigenerazione culturale improntata alla *logofilia*, come instancabile ricerca di senso a fronte di quel nichilismo pratico e *logofobico* – acutamente prefigurato da Sciacca – che ha origine proprio nel processo di assolutizzazione del "relativo", che non può mai essere il fine di se stesso. Ogni prodotto culturale, essendo opera dell'essere umano, porta, ad esso immanente, il problema metafisico della persona che è sproporzione costitutiva, tensionalità infinita verso il mistero dell'essere, che non si lascia intrappolare nella narrazione per immagini ed è irriducibile alla logica del possesso e del consumo.

LA DEMOCRAZIA NELL'ETÀ DELLA RIVOLUZIONE DIGITALE E DELL'ARTIFICIAL INTELLIGENCE

Tommaso Valentini

ABSTRACT

"Fine della democrazia" o comunque un suo radicale indebolimento causato dall'"autoritarismo digitale" o dalla "algocrazia"? In alternativa viene qui presentato il cyberumanesimo, che dell'umanesimo classico riprende il tema della centralità dell'uomo: il digitale è qui visto come uno strumento dell'ingenium che l'uomo deve saper utilizzare per raggiungere obiettivi costruttivi e lo human flourishing. Il cyberumanesimo può essere il presupposto culturale per una riappropriazione critica della democrazia, da parte dei cittadini, nell'età dell'AI.

Parole chiave: algocrazia, cyberumanesimo, democrazia, etica della responsabilità, filosofia politica.

"End of democracy" or, in any case, its radical weakening caused by the "digital authoritarianism" or "algocracy"? As an alternative, cyberhumanism is presented here, which takes up the theme of the centrality of man from classical humanism: digital technology is considered as a tool of the ingenium that man must know how to use to achieve his goals and his human flourishing. Cyberhumanism can be considered as the cultural assumption for a critical reappropriation of democracy by citizens in the digital age.

Keywords: algocracy, cyberhumanism, democracy, ethics of responsibility, political philosophy.

1. *The Digital Age*: mutamenti antropologici e sociali nell'età dell'onlife

La rivoluzione digitale in atto negli ultimi lustri ha portato a cambiamenti repentini che stanno coinvolgendo, e talvolta stravolgendo, in misura sempre maggiore le sfere del lavoro, della comunicazione, del marketing, della sanità pubblica, dell'attività giurisprudenziale, della politica

e quindi della vita privata di ogni individuo che vive nelle società tecnologicamente avanzate. Il processo di informatizzazione, gli sviluppi della *Industry 4.0* con i connessi enormi successi dell'*Artificial Intelligence* (AI) ai quali stiamo assistendo, desta in ognuno meraviglia e spinge a porre nuovi interrogativi sul futuro della nostra società. Autorevoli interpreti del *digital age*, quali Marc Prensky, Paolo Benanti e Paolo Ferri, hanno parlato di una vera e propria trasformazione antropologica in atto, a partire dai cosiddetti "nativi digitali" e dalla "generazione Google": come è noto, i *digital natives* sono i nati dopo la metà degli anni Ottanta, caratterizzati dalla diffusione di massa del PC a interfaccia grafica, mentre i rappresentanti della *Google Generation* prendono il nome dal celebre motore di ricerca informatico entrato in funzione nel 1998: si tratta di persone cresciute in una società multischermo e che considerano le tecnologie come un elemento naturale, senza provare nessun disagio nel manipolarle e nell'interagire con esse¹. Un disagio che spesso avvertono invece gli "immigrati digitali" (*digital immigrants*), ovvero coloro che sono nati prima dell'avvento delle tecnologie digitali e che le hanno adottate in un secondo tempo, costretti da motivi lavorativi e sociali, e che pertanto possono essere classificati quali "apocalittici e integrati", così come faceva notare il semiologo Umberto Eco, in anni in cui la rivoluzione informatica era ancora ai suoi esordi².

Oggi non ha neanche più senso porre la domanda se siamo *on line*, cioè connessi a qualche *digital device*, oppure *offline*, disconnessi dalla sfera del virtuale/digitale: siamo semplicemente *onlife*. Come è noto, il termine *onlife* è un neologismo coniato dal filosofo italiano Luciano Floridi per indicare la trasformazione ontologica in atto e che si può sperimentare quotidianamente: l'attuale, costante condizione di connessione a Internet (*online*) anche nella vita (*life*), ridefinisce un confine

¹ Si vedano anche i volumi di P. FERRI: *La rivoluzione digitale. Comunità, individuo e testo nell'era di Internet*, Mimesis, Milano-Udine 2001; Id., *Nativi digitali*, Mondadori, Milano 2011.

² Cf. U. Eco, *Apocalittici e integrati: comunicazioni di massa e teorie della cultura di massa*, Bompiani, Milano 1964. Sui motivi di "attualità profetica" e sulla *Wirkungsgeschichte* del celebre volume del semiologo si veda anche A.M. LORUSSO (cur.), *50 anni dopo Apocalittici e integrati di Umberto Eco*, DeriveApprodi, Roma 2015.

sempre più labile tra la sfera analogica della realtà in presenza e quella digitale della realtà mediata dalle tecnologie. Non solo siamo sempre connessi, per motivi di lavoro o di *divertissement* personale: ma siamo costantemente reperibili *anywhere*, in qualunque luogo, e *anytime*, in qualsiasi momento. L'*onlife* indica pertanto una nuova categoria ontologica che ridefinisce i rapporti tra il reale e il virtuale, e che comporta un inedito essere-nel-mondo da parte dell'uomo, divenuto ormai *homo digitalis*, la più recente declinazione dell'uomo tecnologico e prometeico che ha caratterizzato la modernità occidentale.

Luciano Floridi per spiegare la fine delle nette e nitide dicotomie tra il reale e il virtuale ha usato spesso nei suoi *papers* e conferenze l'analogia della "società delle mangrovie". Ecco le sue parole in una WebConference tenuta a Lione nel 2018: «[...] Vorrei descrivere la nostra società come la società delle mangrovie. [...] Le mangrovie crescono in un clima meraviglioso dove il fiume (di acqua dolce) incontra il mare (di acqua salata). Ora immaginate di essere in immersione e qualcuno vi chiede: "l'acqua è salata o dolce?". La risposta è che: "Mio caro, non sai dove siamo. Questa è la Società delle Mangrovie. È sia dolce che salata. È acqua salmastra". Quindi immagina che qualcuno ti chieda oggi: "Sei *online* o *offline*?". La risposta è: "Mio caro, non hai idea di dove ti trovi. Siamo in entrambi"». Le mangrovie, ha spiegato Floridi in un'intervista del 2019, «vivono in acqua salmastra, dove quella dei fiumi e quella del mare si incontrano. Un ambiente incomprensibile se lo si guarda con l'ottica dell'acqua dolce o dell'acqua salata. *Onlife* è questo: la nuova esistenza nella quale la barriera fra reale e virtuale è caduta, non c'è più differenza fra "online" e "offline", ma c'è appunto una "onlife": la nostra esistenza, che è ibrida come l'*habitat* delle mangrovie»³.

³ Intervista di Jaime D'Alessandro a Luciano Floridi: "Vi spiego l'era *Onlife*, dove reale e virtuale si (con)fondono", pubblicata sul quotidiano *La Repubblica* il 29 settembre 2019. Tra i numerosi contributi di L. FLORIDI si veda, in particolare, *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford 2014; tr. it. di M. Durante, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, RaffaelloCortina, Milano 2017.

Agli occhi di numerosi interpreti e osservatori critici dell'attuale età tecnocratica, siamo innanzi a un nuovo stadio dell'evoluzione umana, le cui caratteristiche peculiari sarebbero la dimensione del *multitasking*, dell'ipertestualità e dell'interattività. Il pedagogista statunitense Marc Prensky, studioso dei nuovi processi di apprendimento, ha sottolineato che ci stiamo muovendo tutti, sia pur in modo irregolare e con tempistiche diverse, verso un "potenziamento digitale" (*digital enhancement*) che include le attività cognitive e ha implicazioni dirette anche sulle nostre capacità emotive e decisionali⁴. Il potenziamento digitale è l'esito ultimo del potenziamento tecnologico che ha caratterizzato il progresso moderno e che ha reso l'uomo sempre più dominatore della natura, come indicava in maniera paradigmatica, già nel Seicento, Francis Bacon, profeta dell'età tecnocratica: «*knowledge is power*», «*scientia est potentia*»⁵. Oggi quasi tutta la tecnologia è digitale o è supportata da strumenti digitali e questo ha contribuito ad estendere e arricchire le stesse capacità cognitive dell'*homo technologicus* postmoderno. L'esempio più chiaro e indicativo è quello della memoria: la tecnologia digitale migliora enormemente le capacità mnemoniche umane attraverso gli strumenti di acquisizione, archiviazione e restituzione dei dati. Disponiamo di un immenso patrimonio di dati *online*, costantemente fruibile e a nostra disposizione. Inoltre, la raccolta digitale di dati è di supporto alle stesse decisioni umane: abbiamo più elementi per vagliare meglio e verificare, almeno virtualmente, rischi e benefici delle nostre scelte individuali e collettive. *Laptop, database online*, simulazioni tridimensionali virtuali, strumenti collaborativi *online*, palmari e altri *devices* specifici per diversi contesti sono oggi una realtà quotidiana in molte professioni, anche in campi, apparentemente distanti dalle tecnologie,

⁴ Si veda M. PRENSKY, *H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom*, in *Innovate. Journal of Online Education* 5/3 (2009): <https://nsuworks.nova.edu/innovate/vol5/iss3/1>

⁵ Cfr. lo studio, ormai classico, di B. FARRINGTON, *Francis Bacon, Philosopher of Industrial Science*, H. Schuman, New York 1949; tr. it. di S. Cotta, *Francesco Bacone filosofo dell'età industriale*, prefazione di P. Rossi, Einaudi, Torino 1967. L'espressione «*scientia potentia est*» si trova anche in Thomas Hobbes (*De Homine*, X) che di Bacon, in gioventù, era stato segretario.

come la giurisprudenza e le discipline umanistiche, queste ultime sempre più coinvolte dagli sviluppi dell'informatica umanistica ovvero dalle *Digital Humanities* (DH)⁶. La *digital age* ci ha portati verso un progresso ormai irreversibile. Ci troviamo in un punto di non ritorno: nel digitale *vivimus, movemur et sumus*. Siamo innanzi ad un "cambio d'epoca" delle cui ripercussioni occorre essere ben consapevoli. Innanzi alla rivoluzione digitale e ai molti cambiamenti antropologici, sociali e valoriali che essa comporta, occorre elaborare una rinnovata "etica della responsabilità": individuale, comunitaria e planetaria⁷.

In questo contesto tecno-liquido, caratterizzato da una sempre più massiva diffusione del digitale, dove la violazione dei diritti umani, incluso il diritto alla *privacy* e alla tutela dei dati personali, diventa un rischio per i cittadini, lo Stato democratico ha un compito etico-giuridico fondamentale: «In un'epoca dominata dalla digitalizzazione e dall'intelligenza artificiale, è cruciale che lo Stato democratico intervenga per garantire che queste tecnologie siano usate nel rispetto dei diritti e del benessere dei cittadini. La democrazia, pur non essendo perfetta, rappresenta attualmente la migliore garanzia per la protezione delle libertà individuali. Anche se può sembrare lenta e disordinata, la democrazia favorisce lo scambio di idee e assicura il diritto di ogni cittadino di essere rappresentato. In un'epoca in cui la tecnologia sta assumendo un ruolo sempre più predominante, la democrazia può fungere da baluardo, garantendo un utilizzo etico e responsabile delle tecnologie»⁸.

⁶ A questo proposito mi limito a richiamare il saggio critico di C. DOMENELLA, *Human Enhancement e soggetto Post-Umano alla prova delle DH. Come le tecnologie digitali ci trasformano*, in *Umanistica Digitale* 15 (2023).

⁷ Una nuova etica della responsabilità, anche oltre le definizioni di Hans Jonas e in relazione alla rivoluzione digitale, viene proposta, per esempio, da P. BENANTI, *Digital Age. Teoria del cambio d'epoca. Persona, famiglia e società*, San Paolo, Cinisello Balsamo (Mi) 2020; M. BERTOLASO-G. LO STORTO (curr.), *Etica digitale: verità, responsabilità e fiducia nell'era delle macchine intelligenti*, Luiss University Press, Roma 2021; F. FERRO-L. TADDIO (curr.), *Etica e tecnica nell'era digitale*, Mimesis, Milano-Udine 2022; P. BENANTI-S. MAFFETTONE, *Noi e la macchina. Un'etica per l'era digitale*, Luiss University Press, Roma 2024.

⁸ M. CAMISANI CALZOLARI, *Cyberumanesimo*, prefazione di P. Benanti, *Il Sole 24 Ore*, Milano 2024, 38.

2. *Algocracy*: fine della democrazia deliberativa?

La rivoluzione digitale in atto sta comportando mutamenti, sempre più visibili, anche all'interno della sfera del politico e degli stessi ordinamenti democratici. Se nella seconda metà del sec. XX il dibattito politico si è sviluppato, in larga misura, attorno alla questione del rapporto tra lo Stato e il mercato (quanto potere affidare allo Stato e quanta libertà lasciare in mano al mercato), dando origine a soluzioni ordoliberali, neolibériste, ecc., la *magna quaestio* che sta caratterizzando, in maniera sempre maggiore, il sec. XXI è se la vita umana – nella sua dimensione individuale e comunitaria – debba o meno essere sottoposta al controllo di algoritmi elaborati dall'intelligenza artificiale: la grande questione che si pone oggi è, quindi, come conciliare l'*algocracy* (un potenziale dominio degli algoritmi) con i principi di libertà e di autogoverno che costituiscono l'essenza stessa della democrazia. Il futuro della democrazia, il suo rafforzamento o la sua evaporazione dipendono, infatti, dal modo in cui verrà riconfigurata la *governance* dell'AI.

A questo proposito il dibattito politologico attuale è fortemente polarizzato, assumendo, da una parte, tratti quasi messianici (eredi del mito moderno del progresso) e, dall'altra, tratti distopici che prefigurano una "fine della democrazia" o comunque un suo radicale indebolimento causato da un "autoritarismo digitale" (*digital authoritarianism*). In tale scenario distopico la democrazia, intesa come «government of the people, by the people, for the people»⁹, verrebbe sostituita da un impersonale "governo dell'algoritmo" che con le sue enormi capacità predittive sarebbe in grado di elaborare scelte migliori rispetto a quelle che prenderebbe una maggioranza parlamentare di eletti dal popolo. Alcuni politologi si sono spinti a sostenere che la democrazia liberale, così come viene intesa oggi, con le sue istituzioni e procedure, entro il prossimo secolo sarà una realtà obsoleta e verrà sostituita da forme pericolose di *algocracy*.

⁹ A. LINCOLN, *Gettysburg Address*, discorso tenuto il 19 novembre 1863.

In molti contesti sociali – ad es. in ambito economico-finanziario e in ambito medico – i sistemi algoritmici dell'intelligenza artificiale trovano ormai una stabile applicazione, supportando o persino sostituendo i processi decisionali umani: ciò avviene in ragione della capacità dell'AI di elaborare le informazioni e di affrontare compiti cognitivi con una tempistica e una precisione sconosciute alla ragione umana. L'utilizzo degli algoritmi e dell'AI si sta iniziando ad estendere, con un ritmo crescente, anche nell'ambito politico e nei processi decisionali degli ordinamenti democratici¹⁰. A tal proposito mi limito a due esempi emblematici tratti dalle recenti dinamiche politiche del Giappone e della Nuova Zelanda. Nel 2018 alle elezioni a Tama New Town (un distretto della città di Tokio con oltre 150.000) abitanti è stato presentato, per la prima volta, come candidato un *chatbot*: sorprendentemente è stato il terzo candidato più votato, con oltre 4.000 preferenze. Al *chatbot* è stato dato il nome femminile di Michihito Matsuda: questa "candidata digitale" ha offerto un programma elettorale in cui si impegnava a contrastare la corruzione, a dialogare in maniera costruttiva con le altre forze politiche e a fornire «opportunità eque ed equilibrate per tutti»¹¹. Per realizzare questo ambizioso programma politico, Matsuda ha presentato agli elettori il proprio progetto algoritmico basato sul *Machine Learning* (ML), una specificazione dell'intelligenza artificiale che si occupa di creare sistemi che apprendono o migliorano la *performance*

¹⁰ Sull'utilizzo dei sistemi di intelligenza artificiale nei processi decisionali delle democrazie si vedano, in particolare, Y.-C. CHEN-T.C. HSIEH, *Big Data for Digital Government: Opportunities, Challenges and Strategies*, in *International Journal of Public Administration in the Digital Age* 1 (2014) 1-14; J. HÖCHTL-P. PARYCEK-R. SCHÖLLHAMMER, *Big Data in the Policy Cycle: Policy Decision Making in the Digital Era*, in *Journal of Organizational Computing and Electronic Commerce* 26 (2016) 147-169; P.D. KÖNIG, *Citizen Conceptions of Democracy and Support for Artificial Intelligence in Government and Politics*, in *European Journal of Political Research* 62 (2023) 1280-1300.

¹¹ B. COLE, *AI Candidate Promising "Fair and Balanced" Reign Attracts Thousands of Votes in Tokyo Mayoral Election*, in *Newsweek*, 19 aprile 2018: <https://www.newsweek.com/ai-candidate-promising-fair-and-balanced-reign-attracts-thousands-votes-tokyo-892274>; J. LACHLAN, *There's an AI Running for the Mayoral Role of Tama City, Tokyo*, in *Otaquest*, 12 aprile 2018: www.otaquest.com/tama-city-ai-mayor.

in base ai dati che utilizzano. Il *Machine Learning* viene già utilizzato quotidianamente e con profitto, nei paesi informatizzati, quando si interagisce con le banche, si acquista *online* e si utilizzano i *social media*: la candidata virtuale Matsuda ha esortato gli elettori a votare per questa forma efficiente di intelligenza artificiale che garantisce l'eliminazione delle debolezze emotive degli esseri umani, causa principale delle cattive decisioni politiche e dei conflitti di interesse. A sostegno del suo programma elettorale Matsuda ha sottolineato che le passioni umane e le ideologie unilaterali devono essere sostituite dall'analisi scientifica dei dati e da statistiche sugli aspetti positivi delle proposte politiche e delle richieste dei cittadini.

Anche la Nuova Zelanda si è contraddistinta per l'impiego della tecnologia dell'AI nell'attività politica: nel novembre 2017 è stato inaugurato SAM (*Semantic Analysis Machine*), che nel suo *account* di Twitter ha presentato sé stesso come il «your virtual politician on the future». Si tratta di un *chatbot*, ideato dall'imprenditore neozelandese Nick Gerritsen per rispondere in maniera puntuale a questioni come le politiche in materia di alloggi, istruzione, immigrazione, salari ecc. Gerritsen ha affermato di aver pensato a SAM per sanare il divario tra le promesse elettorali dei politici e ciò che invece essi sono in grado di realizzare, un divario che porterebbe i cittadini all'astensione dal voto e ad allontanarsi sempre di più dalla vita politica: attraverso Twitter (il *social network* che nel 2023 ha preso il nome di X), SAM sarebbe sempre direttamente connesso con i cittadini neozelandesi, i quali possono inviare richieste, presentare loro problematiche o condividere riflessioni¹². *De facto* SAM è stato presentato dai suoi creatori come uno strumento informatico in grado di rendere possibile una "democrazia digitale" che sanerebbe la sempre più visibile scissione tra l'istituzione e il cittadino, la lontananza tra il "palazzo" e il "popolo". Ciò che accomuna i due esempi di

¹² Le funzioni e il programma politico di SAM sono stati presentati anche da Walter Langelaar, *co-founder* di SAM – AI Politician (<https://dukakis.org/michael-dukakis-leadership-fellows-and-honorees/walter-langelaar/>); si veda anche il sito <https://www.digital.govt.nz/showcase/sam-meet-your-politician-of-the-future>.

Matsuda e di SAM è la capacità di analisi dell'impatto di potenziali decisioni. Infatti vengono attivate applicazioni dell'intelligenza artificiale in grado di impostare un ragionamento strategico avanzato o di analizzare un set di dati per risolvere problemi complessi. I creatori di tali "macchine intelligenti" sostengono che esse sono in grado di "governare" in maniera migliore rispetto alla mente umana, proprio in virtù della loro capacità predittiva: «they have the power to structure possibilities»¹³.

Come accade già in ambito finanziario e nei settori del *marketing*, anche i governi si affidano sempre più spesso all'AI per la gestione delle politiche pubbliche, al fine di migliorare la qualità dei servizi e offrire al cittadino garanzie di funzionalità ed efficienza. A questo proposito sono stati individuati almeno cinque benefici connessi all'utilizzo dei sistemi di AI nell'ambito delle politiche pubbliche e nella connessa erogazione dei servizi al cittadino. Ecco i cinque principali vantaggi che sono stati rilevati dai sostenitori dell'AI in ambito politico: 1) i dati elaborati dai sistemi informatici possono consentire a chi governa di personalizzare il singolo servizio, per rispondere in maniera più precisa alle esigenze dei singoli utenti; 2) i sistemi di AI aiutano a classificare i servizi al fine di individuare esigenze di specifiche categorie di utenti, che in tal modo possono ricevere aiuti e incentivi; 3) gli algoritmi possono essere utilizzati per fare previsioni e simulazioni su costi, vantaggi e svantaggi di *policies* basate sull'analisi dei dati; 4) l'AI può essere utilizzata per il monitoraggio costante dell'efficienza dei servizi erogati, evitando ritardi, imprecisioni e inefficienze; 5) l'AI è sicuramente un ottimo strumento che aiuta a valutare le politiche pubbliche, *ex ante* ed *ex post*, facendo emergere dall'analisi gli effetti di medio e di lungo periodo dei provvedimenti adottati¹⁴. Questi cinque *benefits* sono ben presenti anche al

¹³ M. ANANNY, *Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness*, in *Science Technology & Human Values* 1 (2016) 97.

¹⁴ Cf. B.W. WIRTZ-J.C. WEYERER-C. GEYER, *Artificial Intelligence and the Public Sector. Applications and Challenges*, in *International Journal of Public Administration* 42 (2019) 596-615; M. KUZIEWSKI-G. MISURACA, *AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-making in Democratic Setting*, in *Telecommunications Policy* 44 (2020): <https://doi.org/10.1016/j.telpol.2020.101976>.

governo di Singapore, uno Stato che è all'avanguardia per la sua *governance* algoritmica; a Singapore è stato, infatti, sviluppato nell'ormai lontano 2004 un programma tutt'ora attivo e che oggi fa ampio uso di AI. Il nome di questo programma governativo è *Risk Assessment and Horizon Scanning* (RAHS). Esso raccoglie dati provenienti dai *databases* pubblici e privati, utili per garantire la sanità pubblica evitando il rischio di epidemie, per prevenire atti terroristici e per prendere decisioni riguardanti le politiche sociali, anche attraverso l'analisi delle opinioni espresse dai cittadini sui *social*. Come spiega il governo di Singapore nel sito istituzionale, «RAHS provides the government with the information they need to “predict, model, and monitor” national security threats: terrorism, disease outbreaks, political and economic crises, and natural disasters»¹⁵.

In futuro la *public sphere* sarà sempre più caratterizzata dalla presenza dei sistemi di AI, i quali stanno trovando precise applicazioni anche nell'ambito giurisprudenziale – si pensi alla giustizia penale, che può utilizzare l'AI per stabilire il rischio di recidiva di un imputato – o all'ambito delle relazioni internazionali, dove gli algoritmi possono essere utilizzati per prevedere, controllare ed evitare potenziali situazioni di conflitto armato. Tuttavia, accanto alle opportunità e ai vantaggi sopra menzionati, l'impiego dell'AI nella sfera politica lascia questioni aperte ed elementi problematici, come non hanno tralasciato di segnalare Nils J. Nilsson, uno dei più rilevanti studiosi del *Machine Learning*¹⁶, e filosofi della politica quali il tedesco Jürgen Habermas e il sudcoreano Byung-Chul Han,

¹⁵ https://www.theglassroom.org/object/singapore_gov-rahs/. Un altro Stato oggi all'avanguardia per l'utilizzo dell'AI nelle scelte di governo e nella pubblica amministrazione è anche la Nuova Zelanda, il cui governo, nel 2018, ha commissionato un importante *Report* su benefici, vantaggi, svantaggi e pericoli di procedure automatizzate, create da *Operational Algorithms*. A tal proposito si vedano: THE DEPARTMENT OF INTERNAL AFFAIRS (cur.), *Algorithm Assessment Report*, ottobre 2018 (<https://www.data.govt.nz/assets/Uploads/Algorithm-Assessment-Report-Oct-2018.pdf>) e il resoconto di J. LIDDICOAT-C. GAVAGHAN, *The Use of Algorithms in the New Zealand Public Sector*, in *New Zealand Law Journal* 1 (2019) 26-27.

¹⁶ Cf. N.J. NILSSON, *The Quest for Artificial Intelligence. A History of Ideas and Achievements*, Cambridge University Press, Cambridge 2009.

e la giurista italiana Lucia Giuditta Sciannella. Quest'ultima ha messo chiaramente in evidenza quattro aporie e problematicità rispetto all'utilizzo esclusivo degli algoritmi e dell'AI per prendere decisioni politiche in nome di una presunta efficienza e velocità che eviterebbero i tempi lunghi della discussione critica parlamentare e democratica. In primo luogo, secondo Schiannella, «partendo dal presupposto che non esiste una nozione condivisa di "buoni risultati", deve richiamarsi che in una democrazia liberale e pluralista, caratterizzata da visioni legittime poste in competizione tra esse, nessuno può pretendere di avere (e di imporre) l'unica nozione – ritenuta "algoritmicamente" buona e corretta – di quale sia il miglior risultato da conseguire. Esistono concezioni contrapposte di ciò che è da ritenersi "buono" come standard sostanziale con cui valutare una decisione politica»¹⁷. In secondo luogo, l'impiego dell'AI in politica può portare a decisioni univoche, scaturite dall'elaborazione di dati che favoriscono una maggioranza a discapito delle minoranze, per cui verrebbero meno gli elementi fondamentali del dibattito e della mediazione. La democrazia liberale e deliberativa deve, invece, essere basata sempre su «un processo di mediazione, in cui la nostra volontà politica può allontanarci dalle nostre preferenze iniziali e in cui gli interessi sono intesi come qualcosa che si scopre attraverso l'interazione comunicativa»¹⁸. Richiamandosi implicitamente a Hannah Arendt e a Habermas, teorici di una "democrazia deliberativa" incentrata sul dialogo e sulla libera prassi argomentativa, Sciannella sottolinea che le decisioni prese seguendo gli algoritmi eliminano *tout court* il fondamento stesso dell'agire politico, ovvero la libertà dei partecipanti al dibattito pubblico:

«La democrazia persegue determinati risultati, garantendo, in ogni fase del processo decisionale, la continua e uguale libertà di tutti i cittadini di prendervi parte. La libertà che contraddistingue, in ogni momento, l'agire politico, deve essere diretta a favorire pertanto reattività e responsabilità, nonché che nessuna posizione di potere domini

¹⁷ L.G. SCIANNELLA, *Intelligenza artificiale, politica e democrazia*, in *DPCE Online* 1 (2022) 337-348: qui 345.

¹⁸ *Ibid.*

permanentemente sugli altri, sullo sfondo di un pluralismo di opinioni e preferenze contrastanti. In questo senso, la democrazia è progettata per regolare il conflitto sociale, lasciando sempre aperte le decisioni da assumere e gli obiettivi da fissare dinanzi a nuovi sviluppi. Dunque, il processo decisionale assume caratteristiche che non sono suscettibili di calcolo e di ottimizzazione»¹⁹.

Non ultima per importanza è un'ulteriore critica, ampiamente condivisibile, rivolta da Sciannella all'algocrazia, ovvero la creazione di nuove marginalità e di nuove forme ingiuste di esclusione, il cosiddetto *digital gap*:

«Bisogna altresì considerare che, in maniera inconsapevole, ciò che gli algoritmi di AI reputano buono per tutti e che sarà l'orizzonte principale di azione nell'assunzione di decisioni e di politiche, rispecchierà esattamente le preferenze, le opinioni e le abitudini di coloro che sono connessi e che, in forza di ciò, rilasciano continuamente dati, avendo un più forte potere economico per acquistare dispositivi di ultima generazione. Ciò aumenterà l'esclusione e le disuguaglianze sociali ed economiche, mettendo a tacere o allontanando i soggetti che non hanno accesso alla Rete e i cui bisogni e desideri non sono intercettati dal sistema di AI»²⁰.

Anche l'anziano Habermas, meditando su potenzialità e rischi della democrazia nell'era del digitale, è tornato a interrogarsi sul nuovo mutamento di struttura della sfera pubblica, dominata ormai da Internet e dai *social media*, nel contesto di un ecosistema mediale che si presenta ibrido, decentrato e reticolare.²¹ Il filosofo tedesco avverte, giustamente, che tali trasformazioni in direzione di una sempre maggiore digitalizzazione e applicazione dell'AI stanno incidendo sul funzionamento deliberativo degli ordinamenti democratici, producendo una frammentazione della sfera pubblica e una polverizzazione delle opinioni generali: i nuovi *media* stanno producendo una frammentazione dei discorsi,

¹⁹ *Ibid.*, 345-346.

²⁰ *Ibid.*, 345 nota.

²¹ J. HABERMAS, *Nuovo mutamento della sfera pubblica e politica deliberativa*, Raffaello Cortina, Milano 2023. Il titolo originario è: *Ein neuer Strukturwandel der Öffentlichkeit und die deliberative Politik*, Suhrkamp Verlag, Berlin 2022.

hanno una finalità essenzialmente commerciale e pertanto si rivelano un ostacolo per l'autentica comunicazione politica. I nuovi *media*, commenta in proposito Marina Calloni,

«avrebbero un impatto distorto sulla comunicazione politica in quanto vengono a creare ristrette *enclave* di dibattito che impediscono una formazione più articolata del consenso secondo argomenti razionalmente motivati e condivisi, [...] allontanando il cittadino dalla focalizzazione su interessi comuni. Le piattaforme inducono altresì una privatizzazione dello spazio per il discorso pubblico, immettendo finalità commerciali nella funzionalità degli algoritmi e riducendo a mercato le preferenze politiche»²².

Dal canto suo – citiamo sempre la lucida sintesi offerta da Marina Calloni – Habermas rileva che «il paradosso della sfera pubblica digitale consiste nel fatto che tanto più è illimitata quanto più privatizza le comunicazioni entro circoli ristretti. Da una “democrazia del pubblico” si passa a una “democrazia dei micropubblici”. Il vero problema della comunicazione politica non riguarda però tanto la produzione di contenuti da parte di singoli cittadini, bensì la costituzione materiale di informazioni prodotte e manipolate dal potere di grandi corporazioni e lobby politiche»²³. Habermas sottolinea quindi che la *digital age* porta nuove gravi problematiche e sfide per la realizzazione della democrazia: siamo infatti anche nell'età proliferazione delle *fake news* e della cosiddetta “postverità”, «per cui una notizia non è vera in sé, ma lo diventa sulla base della ricezione emozionale del pubblico ricevente»²⁴. L'esito delle riflessioni di Habermas non è tuttavia totalmente pessimistico. Pur richiamando i rischi che comportano le nuove forme di propaganda manipolatoria e il diffondersi di nuove forme di controllo sociale dei cittadini che utilizzano gli strumenti digitali (il «capitalismo della sorveglianza»²⁵),

²² M. CALLONI, *Introduzione all'edizione italiana. Ripensare la sfera pubblica*, in HABERMAS, *Nuovo mutamento*, VI-XXXVI: qui XXXIII-XXXIV.

²³ *Ibid.*, xxv.

²⁴ *Ibid.*

²⁵ A questo proposito si veda il volume della sociologa Shoshana Zuboff, che ha suscitato un ampio dibattito anche in Italia: S. ZUBOFF, *The Age of Surveillance Capitalism. The*

Habermas rimane convinto che il digitale e gli stessi processi di AI, se regolamentati dalla politica per finalità etico-sociali, possono rivelarsi anche un'opportunità di crescita democratica. Ed ecco la conclusione alla quale – sottolinea Marina Calloni – giunge il pensatore tedesco:

«L'utilizzo dei nuovi strumenti comunicativi deve essere regolamentato secondo dettami costituzionali che difendano i diritti umani, in modo da evitare abusi e pericoli, come l'evoluzione dell'intelligenza artificiale ci sta indicando. Se i *big data* non sono di per sé un "male", il problema riguarda piuttosto la loro gestione e finalità, ovvero le "domande" che vengono poste e "risolte" mediante algoritmi. I nuovi *media* contengono potenzialità per il benessere collettivo, allorché viene favorito un approccio sociale finalizzato allo sviluppo di un *welfare* che vada incontro alle necessità della popolazione e permetta una politica democratica dei dati, grazie a un migliore rapporto di fruibilità e di trasparenza tra istituzioni e cittadinanza, come da intento delle piattaforme eGov»²⁶.

A conclusioni ben più pessimistiche giunge invece il filosofo di origini sudcoreane Byung-Chul Han, formatosi in Germania e attualmente docente a Berlino. Questi sottolinea che la digitalizzazione del mondo sta facendo degenerare la democrazia in «infocrazia», ovvero «in vite manipolate dalla rete», come indica il titolo di un suo saggio pubblicato nel 2021. Nella *digital era* la politica è costretta a sottomettersi alla logica dei *mass media*, dominata da grandi oligarchie (i proprietari di piattaforme come Facebook, Instagram, X, ecc.) e da un capitalismo che trasforma i cittadini liberi in consumatori da sorvegliare e sottomettere a logiche di mercato. Egli sostiene giustamente che

«*Big Data* e Intelligenza Artificiale consentono al regime dell'informazione di condizionare il nostro comportamento a un livello posto al di sotto della soglia di coscienza. Il regime dell'informazione si appropria di quegli stati preriflessivi, pulsionali, emotivi del comportamento, che precedono le azioni coscienti. La sua psicopolitica basata sui dati s'in

Fight for a Human Future and the New Frontier of Power, Public Affairs, New York 2019; tr. it.: *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Luiss University Press, Roma 2019.

²⁶ CALLONI, *Introduzione all'edizione italiana*, xxxiii.

sinua nel nostro comportamento senza che ci accorgiamo di questa intromissione»²⁷.

Se la democrazia, *ex definitione*, indica che la sovranità appartiene al popolo, il digitale e i suoi possessori, cioè le grandi multinazionali e i proprietari delle piattaforme *social*, finiscono per lasciare al popolo una sovranità solamente apparente e svuotata di significato: il vero «sovrano è colui che dispone delle informazioni in rete»²⁸. Ci troviamo di fronte a una rinnovata forma di "elitismo politico di carattere tecnocratico", una sorta di "elitismo digitale", che decostruisce e mina alla base i fondamenti stessi della democrazia. In questo quadro pessimistico Byung-Chul Han mostra sfiducia anche nei confronti di quella "democrazia digitale" che era stata invece pensata per garantire forme di partecipazione politica più diretta e motivata. L'autore così commenta amaramente:

«La democrazia in tempo reale, che agli albori della digitalizzazione era immaginata come la democrazia del futuro, si rivela una totale illusione. Gli sciami digitali non costituiscono un collettivo responsabile, politicamente attivo. I *follower*, come nuovi sudditi dei *social media*, si lasciano addestrare dai loro *influencer* come bestiame da consumo. Vengono depoliticizzati. La comunicazione sui *social media*, guidata dagli algoritmi, non è libera né democratica. Essa conduce a una nuova minorità. Lo *smartphone* come dispositivo di sottomissione è tutt'altro che un parlamento mobile: esso accelera la degenerazione della sfera pubblica, in quanto pubblicizza instancabilmente la sfera privata come una *vetrina mobile*. Lo *smartphone* produce degli zombie del consumo e della comunicazione, anziché cittadini responsabili»²⁹.

²⁷ B.-C. HAN, *Infokratie*, Msb Matthes & Seitz, Berlin 2021; tr. it.: *Infocrazia. Le nostre vite manipolate dalla rete*, Einaudi, Torino 2023, 15.

²⁸ *Ibid.*, 16.

²⁹ *Ibid.*, 36. Particolarmente suggestive e inquietanti sono le analogie individuate da Byung-Chul Han tra la sfera del religioso e la sfera del digitale, dominata dagli *influencer* dei vari *social media* (You Tube, Instagram, Tik Tok, ecc.). L'atteggiamento dei *follower* ha degli elementi quasi mistici e, in ogni caso, è caratterizzato dall'assenza di pensiero critico: «Gli *influencer* sono adorati come modelli esemplari: così il tutto riceve una dimensione

Secondo il filosofo coreano, quella prodotta dalla Rete e dai sistemi digitali è una «comunicazione senza comunità», una comunicazione di «solitudini interconnesse», incapaci di elaborare criticamente un pensiero dissonante e condiviso, e poco motivate nella riconquista di spazi persi di libertà politica. L'infocrazia produce utenti docili e remissivi, soggetti incapaci di elaborare un pensiero critico e di dividerlo adeguatamente nella sfera pubblica. Byung-Chul Han è quasi drastico nell'affermare che siamo innanzi alla «fine dell'agire comunicativo» e della «democrazia deliberativa» fondata sul sistema dei partiti politici, espressioni di differenti interessi e diverse visioni del bene comune:

«A breve la democrazia partitica non avrà più ragione di esistere: essa cederà il posto all'*infocrazia come post-democrazia digitale*. I politici verranno sostituiti da esperti e informatici, che *amministreranno* la società [...] La politica verrà dissolta in un sistema manageriale basato sui dati. Le decisioni socialmente rilevanti verranno prese attraverso i *Big Data* e l'Intelligenza Artificiale. Ci saranno ancora discorsi politici, ma saranno secondari»³⁰.

Secondo Byung-Chul Han siamo entrati in un mondo dove «la sfera del discorso pubblico è sostituita dall'analisi dei dati per mezzo dell'Intelligenza Artificiale, il che significa la fine della democrazia»³¹. In questo scenario distopico, ma tuttavia realistico, le poche persone proprietarie e amministratrici delle tecnologie digitali più avanzate avranno il controllo su ogni aspetto della vita pubblica e privata dei cittadini, e

religiosa. Gli *influencer*, nella veste di guide motivazionali, si atteggiavano a salvatori. I *follower* partecipano alla loro vita come discepoli, comprando i prodotti che gli *influencer* ingiungono di consumare nella messa in scena della loro quotidianità. Così i *follower* prendono parte a una *eucaristia digitale*. I *social media* somigliano a una chiesa: il *like* è il loro *Amen*. Lo *sharing* è la comunione. Il consumo è la salvezza. La ripetizione come drammaturgia dell'*influencer* non produce noia e *routine*, piuttosto conferisce al tutto il carattere di una liturgia. Al tempo stesso, gli *influencer* presentano i prodotti di consumo come strumenti di auto-realizzazione. Così ci consumiamo fino alla morte, ci realizziamo fino alla morte. Consumo e identità arrivano a coincidere: l'identità stessa diventa una merce» (*ibid.*, 11-12).

³⁰ *Ibid.*, 52-53.

³¹ *Ibid.*, 57.

le aziende private potranno superare in potenza anche gli Stati. Tale scenario distopico presentato da Byung-Chul Han è stato elaborato avendo come punto di riferimento la prospettiva dei cosiddetti "dataisti", cioè di coloro che sostengono il valore esclusivo di una conoscenza basata su dati informatici e sulla loro elaborazione, e che ritengono ormai superflua la produzione umana di teorie e ipotesi scientifiche o filosofiche. L'intento di Byung-Chul Han, nel delineare un ritratto argomentato ma crudo dell'età in cui viviamo, è dunque quello di lanciare un monito agli uomini di oggi che vivono nelle società tecnocratiche e fortemente informatizzate. In tal senso il suo *pamphlet* sull'infocrazia costituisce una sorta di avvertimento, un *warning*: la democrazia è un bene prezioso ma fragile, può essere perso facilmente, soprattutto nell'età dove predominano il digitale e l'AI. Occorre quindi che i cittadini e le istituzioni politiche siano ben consapevoli dei pericoli dovuti a una possibile metamorfosi della democrazia in algocrazia.

In maniera simile a Byung-Chul Han, anche la sociologa statunitense Shoshana Zuboff ha criticato l'immagine fortemente riduttiva dell'essere umano proposta dai "dataisti", difendendo la visione di un'autentica politica democratica basata sullo spirito critico, sulla *parresia* e sulla libertà dei cittadini da ogni forma di controllo (anche digitale). Sostiene giustamente Zuboff che, «Se nei decenni a venire dovremo ridare forza alla democrazia, dovremo ritrovare l'indignazione e il senso di lutto per quello che ci stanno rubando». E ancora: «È in gioco la sovranità dell'individuo sulla propria vita: la possibilità di essere autori della nostra storia. È in gioco l'interiorità che ci dà la volontà di volere e gli spazi pubblici nei quali utilizzarla»³².

³² ZUBOFF, *Il capitalismo della sorveglianza*, 535.

3. Il cyberumanesimo e i presupposti culturali per una rinnovata vita democratica

Innanzi alle sfide che ci presenta l'età digitale, con le connesse e talvolta inquietanti forme di transumanesimo che propongono un'ibridazione tra l'uomo e la macchina, occorre ripensare e riproporre i valori della tradizione umanistica. Si tratta dei valori costitutivi delle società occidentali, quali per esempio la *dignitas personae* (con la sua matrice cristiana) e le istanze di *liberté, égalité, fraternité*, che hanno reso possibile la teorizzazione dei diritti umani e la creazione stessa di istituzioni democratiche³³. In un'epoca in cui l'uomo appare sempre più asservito alle tecnologie di cui egli stesso è il creatore, è necessaria una filosofia che lo renda nuovamente cosciente della sua autonomia e gli fornisca solide basi teoretiche per orientare verso il *bonum commune* le enormi potenzialità del digitale.

In anni recenti i filosofi tedeschi Julian Nida-Rümelin e Nathalie Weidenfeld hanno proposto il concetto di «umanesimo digitale», sviluppato in un importante volume tradotto in più lingue³⁴. I due pensatori propongono un uso consapevole della tecnica, in cui la dimensione dell'*humanum* rimanga sempre al centro: la tecnica è fatta per l'uomo e non viceversa. Essi mettono in discussione alcune pretese esorbitanti dei fautori dell'AI – ad es., che sia possibile una replica artificiale della mente umana – e propongono un'alternativa all'imperante ideologia della Silicon Valley. La loro visione del rapporto uomo/macchina è equilibrata: non intravede scenari apocalittici, perché continua a confidare nella ragione umana e, allo stesso tempo, diffida delle prospettive techno-entusiastiche, perché sa riconoscere i limiti della tecnologia digitale.

³³ Interessanti riflessioni sulle profonde connessioni tra la cultura umanistica e la comparsa delle democrazie liberali in Occidente sono state proposte anche dal francese Philippe Nemo nel volume *Qu'est-ce que l'Occident?*, PUF, Paris 2004, tr. it.: *Che cos'è l'Occidente?*, Rubbettino, Soveria Mannelli (Cz) 2009.

³⁴ J. NIDA-RÜMELIN-N. WEIDENFELD, *Digitaler Humanismus. Eine Ethik für das Zeitalter der Künstlichen Intelligenz*, Piper, München 2018; tr. it.: *Umanesimo digitale. Un'etica per l'epoca dell'intelligenza artificiale*, FrancoAngeli, Milano 2019.

Il *Digital Humanism* è pertanto una prospettiva che pone al centro l'uomo nella sua "unicità irripetibile" (*human uniqueness*) e ribadisce che la funzione di tutte le tecnologie, compresa la sfera del digitale, è quella di aiutare l'uomo ad ampliare le sue capacità. Tuttavia, come rileva anche Marco Camisani Calzolari, i più recenti sviluppi dell'AI rendono ancora più difficile per l'uomo il compito di rimanere «il signore e il custode della *téchne*»:

«Adesso i grandi temi si sono spostati più in alto, su piattaforme interconnesse, tra fisico e algoritmi *online*, a distanza. Una distanza che è più una barriera, a volte impenetrabile, di cui non ci è dato sapere cosa ci sia dietro, cosa sia connesso a chi, di chi siano i dati alla base delle scelte che vengono operate dall'AI, quali siano gli algoritmi alla base di queste e chi li gestisce»³⁵.

Riflettendo in profondità sui possibili usi antiumanistici delle tecnologie digitali, che possono essere asservite a una *libido dominandi* in grado di oltrepassare anche i poteri degli Stati democratici, Camisani Calzolari ha proposto la visione teoretica e antropologica di un *cyberumanesimo*. Come è noto, l'etimologia di *cyber* risale al greco antico *kubernétes*, indicante letteralmente "il timone" e per estensione "colui che governa". Camisani Calzolari osserva che «il concetto di "cyber" cattura l'essenza dell'interazione tra l'uomo e la macchina, mentre la sua origine evidenzia l'importanza del controllo e della gestione. Nel contesto del *cyberumanesimo*, quindi, "cyber" non è solo un prefisso, ma un simbolo della nostra aspirazione a dominare la tecnologia digitale, e a non esserne dominati»³⁶. Mentre il *transumanesimo* appare innamorato della tecnologia e pone in essa speranze quasi salvifiche, intravedendo negli sviluppi della *téchne* il completo superamento di limitazioni fisiche e mentali innate³⁷, il *cyberumanesimo* non propone un'auto-progettazione umanista al di là di ogni limite, ma rimane ben

³⁵ M. CAMISANI CALZOLARI, *Cyberumanesimo*, prefazione di P. Benanti, Il Sole 24 Ore, Milano 2024, 40.

³⁶ *Ibid.*, 41.

³⁷ Sui rapporti e le differenze tra il *transumanesimo* e il *postumanesimo* cf. A. ALLEGRA, *Visioni transumane. Tecnica, salvezza, ideologia*, Orthotes, Napoli 2017; *Id.*, *La persona è antiquata:*

ancorato alla realtà e resta fedele alla vulnerabilità della condizione umana. Insomma, il *cyberumanesimo* «non mira a creare una nuova forma di vita umana, come invece proposto dai transumanisti, e conserva dosi di scetticismo riguardo alle visioni utopiche.

Tuttavia rimane fiducioso nella capacità umana di plasmare le opportunità offerte dal digitale»³⁸. Questa prospettiva si qualifica

«per un approccio olistico che integra l'uomo, il digitale, la macchina e l'intelligenza artificiale. L'uomo non può osservare passivamente i cambiamenti, ma deve agire in modo attivo per influenzare e creare il digitale. È lui che deve guidare le decisioni delle macchine e dell'intelligenza artificiale. Questa integrazione non dovrebbe portare alla limitazione della libertà umana, ma piuttosto alla sua espansione attraverso l'uso responsabile e consapevole delle tecnologie dell'informazione»³⁹.

Il *cyberumanesimo* si presenta, quindi, come una versione rinnovata, *in the digital age*, del concetto di *homo faber*, che affonda le sue radici nella cultura classica e segnatamente nell'umanesimo italiano del Quattrocento: si tratta della visione dell'uomo come creatura libera e creativa, consapevole della sua dignità, delle sue enormi potenzialità artistiche e tecniche, ma anche dei suoi limiti, come indicano i due manifesti più celebri della cultura umanistica italiana: il *De dignitate et excellentia hominis* (1450-1452) di Giannozzo Manetti e l'*Oratio de hominis dignitate* (1486) di Giovanni Pico della Mirandola⁴⁰. Richiamandosi al tema classico della dignità dell'uomo, garantita dal libero arbitrio, dall'essere autonomo e padrone di sé (*compos sui*), il *cyberumanesimo* ribadisce l'importanza cruciale del dominio dell'uomo anche sulla sfera del digitale: il digitale deve venir considerato come uno strumento, un

umanesimo, transumanesimo, postumanesimo, in C. CIANCIO-G. GOISIS-V. POSSENTI-F. TOTARO (curr.), *Persona. Centralità e prospettive*, Mimesis, Milano-Udine 2022, 121-138.

³⁸ CAMISANI CALZOLARI, *Cyberumanesimo*, 45-46.

³⁹ *Ibid.*, 44.

⁴⁰ Sul ritorno alla sapienza umanistica degli antichi per contrastare la colonizzazione della stessa interiorità umana ad opera dei *social media* e dei connessi sistemi di AI cf. M. CRIPPA-G. GIRGENTI, *Umano, poco umano. Esercizi spirituali contro l'intelligenza artificiale*, Piemme, Casale Monferrato (Al) 2024.

alleato, un frutto complesso dell'*ingenium* che l'uomo utilizza per raggiungere i propri obiettivi costruttivi e il proprio *human flourishing*.

Questa prospettiva

«mette in evidenza l'essenzialità di una rete che connette anziché separare e di una tecnologia che serve piuttosto che dominare. Il suo obiettivo non è di ostacolare il progresso tecnologico, bensì di favorire lo sviluppo umano, sfruttando tutte le potenzialità offerte dal digitale per arricchire, ottimizzare e rendere sostenibile la nostra esistenza»⁴¹.

La visione cyberumanista rimane scettica anche nei confronti di una equiparazione ontologica e funzionale tra la mente umana e la robotica più avanzata basata sull'AI. In queste considerazioni il *cyberumanesimo* trova un alleato anche nel linguista americano Noam Chomsky, esperto di processi cognitivi e difensore dell'unicità della mente umana: «La mente umana – ha affermato di recente Chomsky – non è, come la ChatGPT e i suoi simili, una macchina statistica e avida di centinaia di *terabyte* di dati per ottenere la risposta più plausibile a una conversazione o la più probabile a una domanda scientifica». Al contrario, essa è «un sistema sorprendentemente efficiente ed elegante che opera con una quantità limitata di informazioni. Non cerca di estrarre correlazioni di dati, ma si sforza di creare spiegazioni», cioè di comprendere le leggi causali e di distinguere il possibile dall'impossibile. Dunque – continua Chomsky – «smettiamola di chiamarla "intelligenza artificiale" e definiamola per quello che è e fa: un "software di plagio", poiché non crea nulla, ma copia opere esistenti, di artisti esistenti, modificandole abbastanza da sfuggire alle leggi sul *copyright*»⁴².

⁴¹ CAMISANI CALZOLARI, *Cyberumanesimo*, 45.

⁴² N. CHOMSKY, *The False Promise of ChatGPT*, in *New York Times*, 8 marzo 2023: <https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.htm>. Come ha di recente ribadito anche Federico Faggin, il fisico italiano inventore del microprocessore, la coscienza umana è "irriducibile", nel senso che non può essere spiegata completamente in termini di processi fisici o computazionali e pertanto non potrà mai trovare un equivalente nei sistemi di intelligenza artificiale. La coscienza umana è un fenomeno emergente che sorge da complessi processi neurali, e non può essere completamente spiegato da tali processi. A questo proposito si veda il volume che ha sollevato vivaci dibattiti: F. FAGGIN, *Irriducibile. La coscienza, la vita, i computer, la nostra natura*, Mondadori, Milano 2022.

Non possiamo dunque permetterci di essere superficiali o ingenui nell'approccio alle nuove tecnologie: dobbiamo conoscerle, capirle, studiarle, utilizzarle in maniera eticamente corretta e saperle valutare sia nelle loro potenzialità che nei loro limiti. Solo così potremo prevedere e gestire in maniera opportuna e preveggenete i potenziali effetti, anche negativi, che i sistemi di AI possono avere sulle nostre vite, sulla nostra società, sulle nostre democrazie. La prospettiva cyberumanista può essere considerata, pertanto, come il presupposto culturale per una riappropriazione critica della democrazia, da parte dei cittadini, nell'età del digitale. Il *cyberumanesimo* non è solo una filosofia astratta, ma è anche una pratica, un *modus agendi*, un impegno costante e consapevole di conoscenza e di controllo tecnologico, soprattutto laddove i sistemi di AI stanno diventando sempre più utilizzati e pervasivi, come nell'ambito della stessa politica.

In tale contesto, che esige una rinnovata etica della responsabilità comunitaria, tornano di sorprendente attualità le riflessioni ontologiche ed etico-politiche elaborate da Luigi Stefanini negli anni Cinquanta, in un clima di rigenerazione civile, dopo la catastrofe del secondo conflitto mondiale e la caduta dei regimi totalitari, dove l'uomo europeo ha conosciuto gli esiti drammatici della perdita della libertà. Stefanini ribadisce il valore fondamentale della creatività umana, tra i cui prodotti vi è anche la tecnica, chiamata a conservare sempre il suo strettissimo legame ontologico con il suo creatore, cioè l'uomo. Ed ecco così l'enunciato col quale Stefanini riassume il suo personalismo filosofico, che egli stesso intende come un erede e come una specificazione dell'umanesimo classico: «L'essere è personale e tutto ciò che non è personale nell'essere rientra nella produttività della persona, come mezzo di manifestazione della persona e di comunicazione tra le persone»⁴³. Riprendendo e aggiornando questo sfondo teoretico, possiamo dire che anche le nuove tecnologie digitali devono essere sempre considerate come espressione della "produttività umana" e come un mezzo

⁴³ L. STEFANINI, *Personalismo sociale* [1952], Introduzione di A. Rigobello, Studium, Roma, 1979, 7.

per migliorare la vita individuale e sociale, così come auspicato anche dalla più recente prospettiva cyberumanista, attenta a mettere bene in evidenza potenzialità e limiti del digitale.

Mi fa piacere concludere queste riflessioni sul futuro della democrazia nell'età del digitale richiamando alcune parole di Stefanini, tratte dal *Personalismo sociale* (1952), che infondono ottimismo e speranza. Sono parole tese a ricordarci che la vita democratica è per l'uomo «pegno e primizia», è un alto ideale etico-politico da riscoprire costantemente e da valorizzare, un ideale per cui vale sempre la pena impegnarsi e lottare: «L'ordinamento democratico del mondo politico è il più alto e il più ideale, ma è anche il più difficile e il più impegnativo. Esso è un paradigma, il quale pone compiti immani agli uomini che vivono in esso. La democrazia non dev'essere accolta con leggerezza, come un dono, ma come una prova, la più ardua per l'umanità presente»⁴⁴. Per Stefanini la realizzazione di un'autentica democrazia è un compito etico, un ideale regolativo che deve guidare ogni scelta politica, è «un atto di fede che l'umanità fa su sé stessa». Oggi, nell'età della tecnologia digitale, con le opportunità ma anche con i rischi per la stessa democrazia, l'umanità è chiamata a rinnovare questo «atto di fede su sé stessa».

⁴⁴ *Ibid.*, 4-5.

NOTE CONCLUSIVE

Gregorio Piaia

È il momento di tirare le fila di questa giornata così densa di informazioni e spunti di riflessione e il cui tema si potrebbe così sintetizzare: “IA pro persona o IA vs persona”? In altre parole, giocando sulla contrapposizione: “idolatria” della tecnica informatica o “fobia” verso questa tecnica? Se i progressi resi possibili dall’IA appaiono straordinari (basti pensare al settore sanitario o più semplicemente al nostro quotidiano pensare e agire), i pericoli che un uso incontrollato dell’IA rappresentano per la coscienza e il libero arbitrio dell’essere umano in quanto “persona” sono stati qui bene evidenziati e sollevano molte ombre. Nella sua *Introduzione* Roberto Prete ha ricordato come il premio Nobel 2024 per la Fisica sia stato assegnato a due protagonisti dello straordinario sviluppo dell’IA, John Hopfield e Geoffrey Hinton. Un riconoscimento senza dubbio significativo non solo sotto il profilo scientifico ma anche per il ruolo centrale che l’IA sta assumendo nella nostra vita d’ogni giorno. Ed è un ruolo non privo di ombre se si tiene presente che uno dei due premiati, Geoffrey Hinton, dopo essere stato per parecchi anni un dirigente di Google, la potente e onnipresente azienda informatica di cui molti di noi si servono ogni giorno, ha rinunciato a tale incarico «perché – ha dichiarato testualmente – non mi sentivo libero di parlare dei rischi dell’intelligenza artificiale. [...] Me ne sono andato per poter parlare dei suoi pericoli»¹. Una scelta a prima vista sconcertante, ma che bene esprime i dubbi e i timori di un protagonista dello straordinario progresso tecnologico di cui noi siamo solo spettatori e fruitori più o

¹ F. PROVINCIALI, *Il Nobel per l’I.A. a Geoffrey Hinton fa riflettere*, in *Mente politica* 11, n. 76 (19.10.2024), rivista online consultata il 20.10.2024.

meno entusiasti. Ed è una scelta che fa venire in mente le perplessità e il tormento interiore che nel secolo scorso provarono alcuni scienziati di fronte alle possibilità a un tempo esaltanti e terrificanti che si aprivano con l'impiego dell'energia atomica.

IA vs persona, dunque? No! La contrapposizione da cui ho preso le mosse è certamente ad effetto ma è semplicistica e riduttiva: occorre andare oltre questa logica binaria (buono/cattivo), mettendo invece in luce le varie sfaccettature di una realtà così complessa e ambivalente qual è quella compendiata nella sigla IA e alla quale guardiamo con entusiasmo e insieme con paura, con speranza ma anche con timore sulle sorti della "persona". Richiamo allora, in rapida sintesi, alcuni punti a mio avviso essenziali che sono emersi con chiarezza nel corso del convegno. Inizio da una domanda che si avverte sempre più di frequente: un robot di ultima o d'imminente generazione, capace di comunicare con altri robot a nostra insaputa (il *black box* con la sua intrinseca "opacità") e capace di prendere delle decisioni, può essere considerato come un "agente morale", dotato in qualche modo di coscienza e quindi di responsabilità etica e giuridica?

Sì (o quasi...), se teniamo presente che il nostro rapporto con queste macchine particolarmente evolute e "collaborative" non si esaurisce sul puro piano strumentale: non siamo solo "accanto" al robot che ci fornisce informazioni e ci aiuta a risolvere problemi, ma con il robot siamo "interattivi", nel senso che le informazioni e gli input che noi gli diamo e esso ci offre influiscono non solo sulla quantità delle nostre conoscenze ma anche sulla nostra sfera emotivo-affettiva, insomma finiscono con l'orientarci e condizionarci nella nostra vita quotidiana, ovvero nel modo in cui guardiamo a noi stessi e agli altri, e di conseguenza agiamo...

No, se si tiene presente che il nostro cervello non è solo una "macchina" assai perfezionata che interagisce con altre macchine di elevato livello. Questa visione riduttiva del nostro essere-persona è ampiamente diffusa, frutto del razionalismo matematico che si è imposto in età moderna e per il quale solo ciò che nell'uomo è esteriore e misurabile è degno di attenzione da parte dello scienziato, per cui la stessa psicologia, se vuole essere "scienza", deve limitarsi a studiare e misurare

il nostro comportamento esteriore. Già, ma allora Freud dove lo mettiamo? Che fare del “resto” della nostra umanità, del nostro sentirsi felici o infelici, liberi od oppressi, del nostro amare e soffrire, del nostro pregare e imprecare?... Detto in altri termini, nell’uomo le funzioni cerebrali sono “concomitanti” con l’attività della nostra coscienza, ma non “determinanti”. I tempi appaiono maturi per recuperare una visione “integrale” dell’uomo, la cui realtà va ben oltre la macchina più complessa, anche se questa ci batte in fatto di quantità e rapidità delle informazioni.

A questo proposito ho trovato stimolante un libro che è stato menzionato anche nel corso di questo convegno: il titolo è *Oltre l’invisibile. Dove scienza e spiritualità si uniscono* (Mondadori, Milano 2024); l’autore è Federico Faggin, lo scienziato di origine vicentina che inventò il *microchip*, ovvero il circuito integrato miniaturizzato che per la sua elevatissima capacità di memoria costituisce l’elemento base del microprocessore, e poi il *touch-screen*, ossia il *monitor* che a un nostro tocco trasmette comandi a un dispositivo digitale. Nato da una serie di conversazioni dell’autore con Viviana Sardei, già nel titolo il volume suona come una provocazione, giacché parlare oggi di “spiritualità” appare fuori luogo e fuori tempo, in un’epoca in cui un materialismo spicciolo e più o meno inconsapevole sembra caratterizzare la mentalità corrente. Il tema di fondo è stato ripreso dallo stesso Faggin al “Wired Next Fest 2024 Trentino”, svoltosi a Rovereto alla fine dello scorso mese di settembre e dedicato all’innovazione e all’impatto delle tecnologie digitali. L’intervento dello scienziato italo-americano viene così sintetizzato nel resoconto apparso su Internet: «È inutile pretendere che il computer capisca, non capirà mai un tubo!». Una sintesi che non pecca certo di finezza, ma che rende efficacemente il distacco che permane fra la persona, “corpo pensante” e quindi cosciente e responsabile, e la macchina più avveniristica che riesce a imitare modalità prettamente umane, sin quasi a sovrapporsi ad esse o a prenderne il posto. V’è di più: in *Oltre l’invisibile* lo scienziato Faggin si oppone al dilagante scienziatismo che nega l’esistenza della “coscienza” e ripropone – muovendo proprio dalla tecnologia elettronica – quella distinzione/cooperazione

fra materia e spirito che per secoli è stata oggetto di appassionate indagini e discussioni e che per un malinteso ossequio verso la scienza si tende ora a misconoscere e ridurre all'orizzonte del puro materialismo².

Sostenuta da uno scienziato di alto livello, questa posizione non può non fare colpo e induce alla riflessione: è interessante che proprio nel momento in cui il comune cittadino (quello che un tempo era chiamato "l'uomo della strada") sembra cedere alla convinzione che dopo tutto il corpo non è altro che una macchina alla quale, volente o nolente, conviene sottostare a quella macchina assai più perfezionata che è l'IA, una voce autorevole faccia invece presente che l'essere umano non è solo una macchina: è qualcosa di più e di diverso, capace di comprendere e amare l'"altro" oltre che di farlo soffrire e morire, e questo perché è dotato di coscienza e di libero arbitrio: in una parola, l'essere umano è "persona" e non soltanto "macchina". Ben venga allora la macchina pensante, ma il modo di pensare dell'uomo non si esaurisce nella sempre più raffinata razionalità meccanica (basata sulla logica binaria) cui mirano gli informatici. E qui mi viene spontaneo ricordare alcuni celebri "pensieri" di Blaise Pascal (1623-1662), che bene esprimono la ricchezza e la complessità (in apparenza contraddittoria) del nostro "essere persone":

«La grandezza dell'uomo sta in questo: che esso ha coscienza della propria miseria. Una pianta non si conosce miserabile. Conoscere di esser miserabile è, quindi, un segno di miseria, ma in pari tempo, un segno di grandezza. [...] L'uomo è solo una canna, la più fragile della natura, ma una canna che pensa, Non occorre che l'universo intero si armi per annientarlo; un vapore, una goccia d'acqua bastano a ucciderlo. Ma, quand'anche l'universo lo schiacciasse, l'uomo sarebbe pur sempre più nobile di quel che lo uccide, perché sa di morire, e [sa] la superiorità che l'universo ha su di lui; mentre l'universo non ne sa nulla. Tutta la nostra dignità sta, dunque, nel pensiero»³.

² Per un approfondimento di questa posizione si veda G. AMENDOLA, *Dimensioni teologiche dell'approccio panpsichista quantistico-informazionale sulla natura: un confronto tra la visione di Federico Faggin e la teologia cristiana*, in *Scientia et Fides* 11/2 (2023) 107-130.

³ B. PASCAL, *Pensieri*, trad., introd. e note a cura di P. Serini, Oscar Mondadori, Milano 1968, 216-217 (nn. 347 e 397 dell'ed. Brunschwig).

Sono parole di un filosofo, si dirà: certo, ma di un filosofo che era anche un grande matematico, anzi un *enfant prodige* che aveva formulato il cosiddetto teorema di Pascal, basato sulla teoria delle coniche, e che, guarda caso, già negli anni 1642-44 aveva progettato e costruito la prima calcolatrice meccanica, la "Pascalina", lontana antenata del nostro computer. Eppure Blaise Pascal, al pari di Federico Faggin, avvertì l'insufficienza di una risposta puramente scientifica e tecnocratica a una domanda che ci tocca tutte e tutti e che l'IA è oggi capace di misurare ma non di "vivere": che senso ha il mio "essere al mondo"? Una domanda – ecco il punto – cui l'IA più evoluta e abilmente programmata può offrire risposte in maniera surrettizia, ma alla quale non possiamo e non dobbiamo rinunciare a rispondere se vogliamo conservare il nostro *status* di "persona".

Tornando alle relazioni e agli interventi di questo convegno, credo che siano affiorate per lo meno due istanze fra loro complementari. La prima riguarda l'urgenza di una "forte" normativa a livello internazionale che riduca il gap che si è venuto creando fra lo sviluppo tumultuoso delle applicazioni informatiche e l'ordinamento giuridico vigente nei singoli Stati, e ponga quindi dei limiti a un liberismo sfacciato e travolgente che finirebbe col consegnare le sorti dell'umano consorzio nelle mani di poche compagnie transnazionali, versione moderna del *deus ex machina*. Le iniziative dirette a una maggiore trasparenza e a un maggiore controllo non mancano, dall'*AI Act* 2024/1689 della Comunità europea all'*Executive Order* emanato negli Usa dal presidente Biden; ma, vuoi per ragioni burocratiche vuoi per gli ostacoli frapposti dai grandi complessi economici nonché militari, queste iniziative rischiano di avere un'applicazione lenta e tardiva, risultando infine superate e quindi inefficaci (insomma, per citare un vecchio proverbio contadino, è come "chiudere la porta della stalla quando la vacca è già scappata" ...).

Questa è però soltanto una faccia della medaglia, di fronte alla quale il comune cittadino si sente soltanto spettatore (o meglio vittima impotente), data la sua lontananza dai centri effettivi del potere. V'è tuttavia un'altra istanza, che ci coinvolge tutti e ciascuno e che tocca il rapporto delicato ma vitale fra "tecnocrazia" e "democrazia": non

basta delegare al potere politico e all'apparato legislativo-giudiziario il compito di fronteggiare l'azione invasiva e pervasiva che l'IA è in grado di compiere. Occorre invece mobilitarsi a tutti i livelli, a partire da quello educativo-formativo ad ampio raggio (scuola e *media*), per rendere consapevoli le ultime generazioni degli effetti alienanti che l'uso ingenuo e indiscriminato dei prodotti informatici produce nel nostro rapporto con la realtà sia naturale che sociale. Occorre far comprendere ai giovani (e ai meno giovani) che il razionalismo scientifico non risolve la complessità e la ricchezza della dimensione umana, perché «scienza senza coscienza non è che rovina dell'anima», come proclamava il buon Rabelais agli albori della modernità. Occorre poi essere consapevoli che, se sul piano individuale ognuno di noi pesa ben poco o niente, è però possibile operare insieme, "fare massa" nel senso positivo del termine, per porre all'attenzione di chi detiene il potere politico ed economico gli effetti di un uso perverso dell'IA che finirebbe col sostituire quest'ultima all'umano. Insomma, far interagire la democrazia con la tecnocrazia, altrimenti il divario si farà sempre più profondo. È questa la sfida che tutti noi dobbiamo affrontare, dal momento che, come ho rilevato nella *Prefazione*, «il futuro non è più quello di una volta».

Profili degli autori

SALVATORE AMATO: professore ordinario di Filosofia del diritto nel Dipartimento di Giurisprudenza dell'Università di Catania, dove insegna anche Biogiuridica, Diritto e letteratura, Intelligenza artificiale e diritto. Membro del Comitato nazionale per la bioetica dal 2001 al 2022. Ha pubblicato oltre duecento saggi su importanti riviste italiane e straniere. Monografie più recenti. *Eutanasie* (2015), *Dentro il diritto* (2016), *Biodiritto 4.0.* (2020).

ANDREA GALLUZZI: ingegnere elettronico-informatico, è anche laureato in Filosofia all'Università di Perugia e in Ontologia trinitaria all'Istituto universitario "Sophia" di Loppiano (Figline e Incisa Valdarno), ove attualmente insegna Filosofia della tecnica. Ha ricoperto vari ruoli nei settori *Information and Communication Technologies* e sviluppo software.

ALBERTO M. GAMBINO: professore ordinario e prorettore vicario dell'Università Europea di Roma, Commissario italiano della European Commission against Racism and Intolerance (ECRI) del Consiglio d'Europa, nel Gruppo di Lavoro "Tecnologie emergenti" del Comitato Nazionale di Bioetica, Presidenza del Consiglio; nella Commissione speciale "Intelligenza artificiale e diritto d'autore" del Comitato Consultivo Permanente per il Diritto d'Autore, Ministero della Cultura, Presidente dell'Italian Academy of the Internet Code.

MARKUS KRIENKE: professore ordinario di Filosofia moderna ed Etica sociale presso la Facoltà di Teologia di Lugano e direttore della Cattedra Rosmini. Inoltre insegna Antropologia filosofica alla Pontificia Università Lateranense e Dottrina sociale della Chiesa alla Facoltà teologica dell'Italia Settentrionale (Milano). È membro dell'Accademia Europea delle Scienze e delle Arti e del Comitato scientifico italiano della Fondazione "Konrad Adenauer".

PAOLO MORO: avvocato abilitato al patrocinio avanti alla Corte di Cassazione e professore ordinario di Filosofia del Diritto e Informatica giuridica all'Università di Padova, ove è presidente del Corso di Laurea in Giurisprudenza 2.0 (sede di Treviso). Dirige la rivista *Journal of Ethics and Legal Technologies* e ha pubblicato vari libri, articoli e contributi su retorica forense, informatica giuridica e tecnodiritto, nonché su didattica giuridica e diritto dello sport.

GREGORIO PIAIA: professore emerito di Storia della filosofia all'Università di Padova, è presidente onorario della Fondazione "Luigi Stefanini" di Treviso.

ROBERTO PRETE: avvocato, è vice-presidente della sezione di Treviso della UGCI (Unione Giuristi Cattolici Italiani).

LEOPOLDO SANDONÀ: docente e direttore presso l'ISSR di Vicenza. È responsabile del progetto Etica e Medicina della Fondazione Lanza e segretario scientifico del Comitato Regionale per la Bioetica (Regione del Veneto). I suoi interessi di

ricerca si concentrano sul pensiero dialogico contemporaneo, sul crinale filosofico-teologico, e sulle implicazioni pratiche specie in ambito bioetico. Autore di numerose pubblicazioni, è stato relatore a convegni di carattere nazionale e internazionale.

FLAVIA SILLI: docente di Introduzione alla storia della filosofia e di Filosofia della Conoscenza alla Pontificia Università Lateranense dove svolge anche corsi sul Personalismo e sull'Estetica personalista del Novecento. È cultore della materia di Storia della Pedagogia all'Università La Sapienza e docente invitato di filosofia della Religione alla Pontificia Facoltà di Scienze dell'Educazione dell'"Auxilium". Fa parte del Comitato direttivo di Persona al centro, Associazione per la filosofia della persona e del Comitato scientifico della Fondazione Luigi Stefanini di Treviso.

LUCIA STEFANUTTI: nata a Treviso, laureata a Ca' Foscari di Venezia in Storia della Filosofia Contemporanea, ha insegnato nei licei. Già nel direttivo dell'associazione Persona al Centro, coordinatrice scientifica e presidente della Fondazione Luigi Stefanini.

LUIGI VERO TARCA: professore onorario presso l'Università Ca' Foscari di Venezia, dove è stato professore ordinario di Filosofia teoretica, direttore del Cestudir (Centro Studi sui Diritti Umani), delegato alla Ricerca scientifica del Dipartimento di Filosofia e Beni culturali e presidente del Corso di laurea in Filosofia. Ha elaborato una proposta filosofica basata sulla "pura differenza", a partire dalla quale ha ideato e promosso numerose attività di

“pratiche filosofiche”, interne ed esterne al mondo universitario.

FRANCESCO TOTARO: già professore ordinario di Filosofia morale e prorettore all'Università di Macerata, dal 2017 al 2024 è stato Presidente del Comitato Scientifico del Centro di Studi Filosofici di Gallarate, con sede a Roma. Ha intrecciato la sua riflessione ontologico-metafisica (*Assoluto e relativo. L'essere e il suo accadere per noi*, 2013) con la filosofia del lavoro (da *Non di solo lavoro*, 1998 a *Lavoro, ozio, festa: riequilibrare l'umano*, 2024) e con i temi dell'economia e della politica.

TOMMASO VALENTINI: professore associato di “Filosofia politica” presso l'Università degli Studi “Guglielmo Marconi” (Roma), dove ricopre anche l'incarico di Vicedirettore del Dipartimento di Scienze umane. Dal 2012 è docente incaricato di “Ermeneutica filosofica” presso la Pontificia Università Antonianum. Dal 2016 è nella direzione scientifica di «Areté. International Journal of Philosophy, Human & Social Sciences». È membro del comitato scientifico della “Fondazione Luigi Stefanini” e del comitato fondatore di “Persona al Centro. Associazione per la Filosofia della persona”. Tra le sue pubblicazioni: *I fondamenti della libertà in J.G. Fichte. Studi sul primato del pratico*, Editori Riuniti University Press, Roma 2012; *Filosofia e cristianesimo nell'Italia del Novecento*, Drengo Edizioni, Roma 2012; è stato curatore (con Flavia Silli) del volume *Il personalismo sociale di Luigi Stefanini. Fondamenti, valori e motivi di attualità*, Studium, Roma 2023.

PUBBLICAZIONI DIGITALI – Triveneto Theology Press

FACOLTÀ TEOLOGICA IN RETE

SEDE A PADOVA



Ciclo Istituzionale

Ciclo di Specializzazione (Teologia pastorale e Teologia Spirituale)

Ciclo di Dottorato

ISTITUTI TEOLOGICI AFFILIATI (Ita)



Concordia-Pordenone – Studio teologico “Card. Celso Costantini”



Gorizia-Trieste-Udine – Studio teologico interdiocesano



Belluno-Feltre-Treviso-Vittorio Veneto – Studio teologico interdiocesano



Verona – Studio teologico “San Zeno”

ISTITUTI SUPERIORI DI SCIENZE RELIGIOSE (Issr)



ISSR “Giovanni Paolo I” di Belluno-Feltre, Treviso, Vittorio Veneto



ISSR di Bolzano-Bressanone



ISSR “Santi Ermagora e Fortunato” di Gorizia, Trieste, Udine



ISSR di Padova



ISSR “Romano Guardini” di Trento



ISSR “San Pietro martire” di Verona



ISSR “Mons. Arnoldo Onisto” di Vicenza



TRIVENETO
THEOLOGY
PRESS

PUBBLICAZIONI DIGITALI OPEN ACCESS
DELLA FACOLTÀ TEOLOGICA DEL TRIVENETO

COLLANA THEOLOGY

- 1 - LORENZO VOLTOLIN (cur.), *Sulla fragilità dell'esistenza. Dialoghi*, 2023, pagine 84, isbn 979-12-81328-01-3.
- 2 - ANTONIO BERTAZZO (cur.), *Quo vadis? Cammino, paradigma per Dio e per l'uomo*, 2023, pagine 68, isbn 979-12-81328-00-6.
- 3 - FRANCESCO PESCE (cur.), *Azione, esperienza, fede. Una prospettiva di teologia pratica*, 2023, pagine 193, isbn 979-12-81328-03-7.
- 4 - LIVIO TONELLO (cur.), *Umanesimo digitale. Educarsi al digitale per educare*, 2023, pagine 137, isbn 979-12-81328-05-1.
- 5 - FRANCESCO PESCE, *Sulla frontiera. Per una teologia alla luce di Evangelii gaudium*, pubblicato il 24.11.2023, nel 10° anniversario di *Evangelii gaudium*, pagine 53, isbn 979-12-81328-07-5.
- 6 - LUIGI SARTORI, *Trittico ecumenico*, a cura di M. Milani, pubblicato nel centenario della nascita di Luigi Sartori, 2024, pagine 66, isbn 979-12-81328-04-4.
- 7 - ANTONIO BERTAZZO (cur.), *La possibile via. Meditazione e preghiera*, 2024, pagine 151, isbn 979-12-81328-10-5.
- 8 - GIUSEPPE CARAMAZZA, *The Gospel of Transformation. Faith, Justice, and the Path to a Renewed Society*, in lavorazione

COLLANA EDUCATION

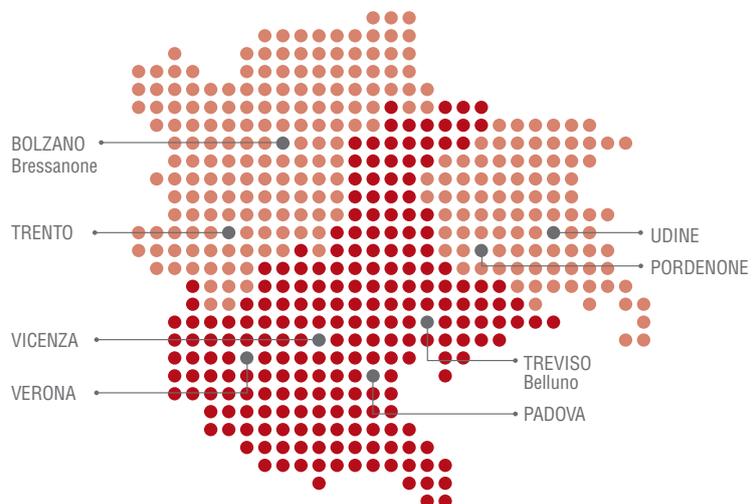
- 1 - ALESSIA FANT, *Scuola e casa comune. Educare con gioia piccoli custodi del creato*, 2023, pagine 85, 2023, isbn 979-12-81328-02-0.

- 2 - LAURA DE FAVERI, *Perdere tempo per guadagnare tempo. La sfida educativa proposta dalla pedagogia della lentezza*, 2023, pagine 84, isbn 979-12-81328-06-8.
- 3 - CHIARA BRUNELLO, *I potenziale religioso del bambino. Percorsi di IRC ispirati a Sofia Cavalletti*, prefazione di Michele Visentin, 2024, pagine 163, isbn 979-12-81328-08-2.
- 4 - MARTINA FAGHERAZZI, *Insegnamento della religione e autismi. Orientamenti pedagogici e didattici per la Scuola dell'Infanzia e Primaria*, prefazione di Cosima Damiana D'Urso, pagine 105, isbn 979-12-81328-09-9.

COLLANA PHILOSOPHY

- 1 - GREGORIO PIAIA-ROBERTO PRETE-LUCIA STEFANUTTI (curr.), *Intelligenza artificiale e tutela della persona umana. Implicazioni etico-giuridiche. Atti del Convegno di studio (Treviso, 20 ottobre 2023)*, 2024, pagine 148, isbn 979-12-81328-08-2.
- 2 - GREGORIO PIAIA-ROBERTO PRETE-LUCIA STEFANUTTI (curr.), *Intelligenza artificiale. Sviluppi futuri e tutela della persona. Atti del Convegno di studio (Treviso, 11 ottobre 2024)*, 2024, pagine 186, isbn 979-12-81328-12-9.

COLLANA BIBLE



Triveneto Theology Press

Facoltà Teologica del Triveneto

Via del Seminario, 7 - 35122 Padova

tel. +39 049 664116

segreteria@fttr.it

www.fttr.it

@facoltateologicatriveneto

@fac_teol_triven



9 791281 328129