

PHILOSOPHY

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)



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 979-12-81328-08-2

Editing: Facoltà Teologica del Triveneto, a cura di Caterina Benvenuti
editoria@fttr.it

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



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)

PHILOSOPHY - 1

RELIGIOUS STUDIES

ABSTRACT

Sviluppo e diffusione dell'intelligenza artificiale e tutela della persona umana: implicazioni etico-giuridiche. Questo il titolo del Convegno svoltosi a Treviso, il 20 ottobre 2023, con il patrocinio dell'Ordine degli Avvocati e l'organizzazione della Fondazione Stefanini, in collaborazione con la sezione trevigiana dell'Unione Giuristi Cattolici. Queste istituzioni sono impegnate da alcuni anni a Treviso nell'organizzazione di convegni e dibattiti su tematiche inerenti alla difesa dei diritti della Persona umana.

La motivazione di fondo del Convegno del 20-10-2023 era quello di comprendere meglio la sfida dell'IA, attraverso un confronto fra esperti di varie discipline, accomunati dall'aver compiuto recenti aggiornati studi sull'Intelligenza Artificiale e da un approccio non pregiudiziale sull'argomento. Le discipline coinvolte spaziano dall'antropologia all'informatica, alla filosofia politica, dall'etica al diritto, dalla linguistica fino alla metafisica.

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

Development and dissemination of artificial intelligence and protection of the human person: ethical-legal implications. *This is the title of the conference organized in Treviso on October 20, 2023, under the patronage of the Lawyers' Order and organized by the Stefanini Foundation in collaboration with the UGCI of Treviso. These institutions are involved in organizing meetings and debates on the defense of the rights of the human person.*

The underlying motivation of the conference was to gain a better understanding of A.I. through a comparison among scholars from various study subjects, ranging from Anthropology to Computer Science, Political Philosophy, Ethics, Law, Linguistics, to Metaphysics.

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
Girolami Maurizio (*sede Fttr*) – responsabile sezione Episteme
Osto Giulio (*Issr di Padova*) – responsabile sezione Praxis

Membri

Barcaro Marco (*Sede Fttr*)
Boscolo Gastone (*Sede Fttr*)
Curzel Chiara (*Issr di Trento*)
Dal Pozzolo Alessio (*Issr di Vicenza*)
Didonè Stefano (*direttore "Studia patavina"*)
Frausin Sergio (*Issr di Gorizia, Trieste, Udine*)
Merlo Luca (*Ita di Verona*)
Toniolo Andrea (*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.

Fondazione Luigi Stefanini

La *Fondazione Luigi Stefanini* è stata costituita a Treviso, città natale del filosofo, il 21 giugno 1996, per volontà dei figli Lucia e Paolo, e di tre appassionati cultori trevigiani di filosofia: Giacomo Bernardi, Gian Paolo Cellini e Renato Pagotto. Scopo della *Fondazione*, assente ogni fine di lucro, e gestita in puro volontariato, è lo studio e l'interpretazione dell'opera di Luigi Stefanini (Treviso 1891-Padova 1956), come pure la diffusione e la riproposta del suo pensiero per lo spessore speculativo e l'attualità che lo caratterizzano. Essa è per questa ragione a disposizione di chiunque sia interessato, per motivo di studio, di ricerca, di lavoro, di cultura, alla sua speculazione filosofica.

Stefanini, illustre docente dell'Università di Padova, tra i maggiori interpreti di Platone, fu il teorico dell'"*imuginismo filosofico*", e soprattutto il metafisico più coerente in campo europeo del movimento personalistico, in diretto confronto con tutte 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 "*Biblioteca Luigi Stefanini*", donata alla *Fondazione* dai figli Lucia e Paolo, ricca di circa duemila volumi, è conservata dalla *Fondazione* stessa presso il Seminario Vescovile di Treviso. Vi si trovano tutte le opere scritte dal Filosofo trevigiano (la maggior parte in edizione originale, talvolta con annotazioni autografe), i volumi della sua biblioteca filosofica personale, nonché gran parte delle opere scritte su di lui.

SOMMARIO

<i>Roberto Prete</i> Premessa	11
<i>Gregorio Piaia</i> Introduzione	15
<i>Vittorio Possenti</i> La persona dinanzi all'ambigua polivalenza dell'IA	17
<i>Riccardo Pozzo</i> <i>Domain Specific Machine Learning</i> nelle scienze umane	35
<i>Andrea Galluzzi</i> Intelligenza Artificiale come "tecnologia del sé"	45
<i>Markus Krienke</i> Intelligenza Artificiale e libertà. Come contrastare la sfida transumana	63
<i>Luigi Vero Tarca</i> MagIA! L'Intelligenza Artificiale e le nuove forme dello spirito	85
<i>Paolo Moro</i> Intelligenza Artificiale per legge. Origini e sviluppi del tecnodiritto contemporaneo	107

Sommario

<i>Leopoldo Sandonà</i> "Luoghi etici" e competenze etiche nell'epoca dell'Intelligenza Artificiale	123
<i>Giuseppe Goisis</i> La sfida del "tecnoumano"	137
<i>Gregorio Piaia</i> Note conclusive	141
Profili degli autori	143

PREMESSA

Roberto Prete

Insieme con la Fondazione “Luigi Stefanini”, la Sezione trevigiana dell’Unione Giuristi Cattolici Italiani ha ritenuto opportuno organizzare questo Convegno per studiare e valutare la diffusione dell’Intelligenza Artificiale (IA) nei diversi settori della società odierna. Lo sviluppo delle tecnologie digitali comporta effetti economici, sociali e culturali di grande rilievo. Nella nostra società queste tecnologie sono talmente sofisticate che stanno sostituendo la persona umana anche nelle funzioni che riguardano la sua entità e coscienza. Si pensa prevalentemente ai vantaggi della tecnologia, sottovalutando i rischi che può comportare. Ci sono invece concreti interrogativi sull’organizzazione dell’economia, sul modo di lavorare, di comunicare, di vivere e di relazionarsi con gli altri. Se queste prospettive risultano male gestite, in assenza di regole adeguate si possono avere conseguenze negative per la dignità e per i valori fondanti della persona, con riferimento ai doveri inderogabili e ai diritti inviolabili costituzionalmente garantiti dalla nostra Carta costituzionale.

È al riguardo significativo che Giovanni Maria Flick, già presidente della Corte costituzionale nonché Ministro della Giustizia e professore emerito di diritto penale alla LUISS di Roma, insieme con la figlia Caterina Flick abbia pubblicato di recente un libro dal titolo *Il mito dell’informatica. L’algoritmo d’oro e la Torre di Babele* (Milano, Baldini & Castoldi, 2022). Richiamandosi al racconto biblico della Torre di Babele, dove gli uomini, per la loro sete e bramosia di conoscenza, ebbero la punizione dell’incomunicabilità nel linguaggio, Flick ha messo in risalto la necessità di trovare quanto prima un equilibrio tra la civiltà degli uomini e la civiltà delle macchine, o meglio la civiltà digitale, richiamandosi anzitutto all’art. 9 della Costituzione, dove si afferma che

«la Repubblica promuove lo sviluppo della cultura e la ricerca scientifica e tecnica. Tutela il paesaggio ed il patrimonio storico e artistico della Nazione. Tutela l'ambiente, la biodiversità e gli ecosistemi, anche nell'interesse delle future generazioni. La legge dello Stato disciplina i modi e le forme di tutela degli animali». Ciò significa rimettere al centro l'essere umano e la sua personalità, garantendo la pari dignità sociale, come recita l'art. 3 della Costituzione («Tutti i cittadini hanno pari dignità sociale e sono uguali davanti alla legge, senza distinzione di sesso, di razza, di lingua, di religione, di opinioni politiche, di condizioni personali e sociali»).

In realtà i timori legati alla diffusione delle tecnologie digitali sono concreti, a partire dal rischio di una perdita massiccia di posti di lavoro. Negli Stati Uniti trecentocinquanta fra studiosi e operatori dell'IA hanno firmato nel 2022 una dichiarazione sul rischio dell'estinzione del genere umano provocata dalle tecnologie digitali. In particolare lo scienziato Geoffrey Hinton, già esperto e operatore di Google, ha sottolineato i rischi che la tecnologia digitale sfugga al controllo umano. Si sono poi concretizzati di recente i rischi per il rispetto del diritto d'autore, tant'è vero che negli USA si è svolto uno sciopero di scrittori, attori e registi che lavorano presso gli *Studios* di Los Angeles, durato per mesi fino al raggiungimento di un accordo.

Non sono mancati i tentativi di istituire delle regole. In Italia c'è stata una pronuncia della Corte di Cassazione in tema di rispetto del diritto d'autore (Cass. Ordinanza 16 gennaio 2023 n. 167) in una controversia tra un autore televisivo e la RAI. Sempre in Italia è stato posto il problema di introdurre l'IA anche nel processo penale, per rendere i verbali degli interrogatori degli indagati e degli imputati molto più celeri e immediatamente trascrivibili, ma al momento nulla è stato deciso. In data 14 giugno 2023 il Parlamento Europeo ha approvato il Progetto normativo della Commissione Europea circa la sua posizione in materia con l'*Artificial Intelligence Act*, una normativa incentrata soprattutto sulla vigilanza per la protezione dei dati personali e sulla garanzia del libero mercato. Ma il problema del rispetto dei diritti umani rimane ancora

Premessa

aperto. In Italia il Garante per la protezione dei dati personali ha preso il Provvedimento n. 114 del 11 aprile 2023 circa l'applicazione e l'uso della ChatGPT. La Fondazione Europea degli Avvocati ha pubblicato recentemente delle linee guida per l'uso dell'IA nell'ambito professionale. Manca però ancora una regolamentazione precisa per l'applicazione di questa innovazione nella legislazione dei singoli Stati membri della Comunità Europea, compresa l'Italia.

INTRODUZIONE

Gregorio Piaia

Credo che in questo periodo, a parte le drammatiche notizie sul conflitto fra i palestinesi di Hamas e Israele, il tema dell'Intelligenza Artificiale (IA) sia tra i più presenti nelle cronache dei quotidiani, ove si susseguono le iniziative (seminari, convegni e dibattiti) sui vari aspetti di una "cosa" che in tempi brevi inciderà in modo radicale sul nostro modo di vivere, lavorare, pensare e comunicare. Proprio ieri (19 ottobre) il XIII Seminario di tecnologie dell'informazione, svoltosi all'Accademia Galileiana di Padova, verteva su "Reti complesse e sistemi dinamici per le Neuroscienze: nuove opportunità e prospettive per la diagnosi e il trattamento personalizzato", e l'ultima relazione era su "L'IA applicata al *neuroimaging*". Il 17 ottobre, sempre a Padova, s'è discusso al Palazzo del Bo su "Brevetti, software e intelligenza artificiale", mentre al IX Festival della dignità umana (Borgomanero, 1-13 ottobre) ci si è posta la domanda "Intelligenza artificiale: siamo pronti al postumano?". A Roma nello scorso settembre il Ministero della Cultura ha organizzato un convegno di ampio respiro su "Intelligenza artificiale: creatività, etica, diritto e mercato", mentre alla LUISS "Guido Carli" – sempre a Roma – è stato affrontato il tema specifico "Intelligenza artificiale, proprietà intellettuale e concorrenza: evoluzioni normative e *Artificial Intelligence Act*". Non solo: giorni fa leggevo su un quotidiano *on line* che con un versamento di Euro 250 si può entrare agevolmente nel giro delle quotazioni del petrolio, gestite da un sistema IA che in brevissimo tempo garantisce un'ottima resa senza fatica alcuna...

Dalla medicina al mercato, dunque; dai sensori in grado di leggere e modificare l'attività cardiaca e cerebrale alla rivoluzione nel campo dei brevetti, fino alla stesura tramite l'IA di romanzi e poesie. E qui si può ricordare come già nel 1971 (cinquant'anni prima di Chat GPT!) lo

scrittore Primo Levi avesse inventato, per un telefilm della RAI, il cosiddetto "versificatore", ossia una macchina capace di creare poesie su indicazioni dell'uomo. Come dire: siamo passati nell'arco di mezzo secolo dalla pura fantascienza alla realtà quotidiana. A questo punto, per una sorta di cortocircuito, mi vengono in mente le sofferte parole con cui nei giorni scorsi lo scrittore Ferdinando Camon ha chiuso un suo corsivo, chiedendosi, alla luce degli orrori e delle bestialità di cui l'uomo è capace, se non sia un male l'estinzione dei leoni e degli orsi, e se non sia invece un bene che l'umanità sparisca dalla faccia della terra. Un appassionato di IA potrebbe rispondere che la soluzione a questo dilemma è a portata di mano: basta introdurre nel cervello un microchip ben congegnato per evitare da parte nostra comportamenti disumani. Certo, ma allora che senso ha parlare di *libertà* e *responsabilità* della persona? E chi mi assicura che l'inserimento in massa dei microchip non sia programmato e gestito per altri fini da un potere più o meno occulto, e che l'IA non finisca col determinare il nostro comportamento, rendendoci simili ad automi? Daccapo, dalla fantascienza alla realtà...

Svariati e talora assai lontani fra loro sono dunque i campi di applicazione dell'IA, anche se legati da una comune base tecnico-scientifica. In questa sede l'intento è di richiamare l'attenzione sulle problematiche etico-giuridiche che sorgono dall'impiego dell'IA e che fanno riferimento alla nozione di *persona* cui si ispirano sia la Fondazione "Luigi Stefanini" sia l'Unione Giuristi Cattolici Italiani, e che va ben oltre la grigia e asettica nozione di *individuo* o *soggetto*. E qui, entro il quadro generale dei rapporti fra tecnica ed etica, e quindi fra tecnologia e diritto, balzano in primo piano i concetti di *libertà* e di *responsabilità* cui s'è appena accennato, con la conseguente esigenza di regolamentare un ambito assai vasto e seducente, ma anche infido.

LA PERSONA DINANZI ALL'AMBIGUA POLIVALENZA DELL'AI

Vittorio Possenti

Abstract

Il contributo argomenta l'impossibilità che l'IA sia un'intelligenza a pieno titolo. Dal lato etico e antropologico occorre un'attenta cautela perché spesso si presuppone che l'aumento della potenza sia un bene in sé. I *Cloud Empires*, ossia le grandi piattaforme che possono per alcuni aspetti assomigliare agli Stati, se ne differenziano però profondamente perché non debbono rendere conto del loro operato ai cittadini: non vi è una "democrazia digitale". Vi è timore dinanzi all'IA, perché gli esseri umani potranno usarne male, infliggendosi gravi danni.

Parole chiave: Intelligenza umana, Persona, Etica dell'IA, Regolamentazione mondiale, Rischio dei Clouds Empires

The contribution argues the impossibility that the AI is an "intelligence". From the ethical and anthropological side, careful caution is needed because it is often assumed that the increase in power is a good in itself. The Cloud Empires, i.e. the large platforms that may in some respects resemble States, however differ profoundly because they do not have to account for their actions to citizens: there is not a "digital democracy". There is fear in front of AI, because human beings could be able to misuse it, inflicting serious damage on themselves.

Keywords: Human intelligence, Person, AI ethics, Global regulation, Risk of Cloud Empires

Prologo

La questione dell'IA appare di grande difficoltà in rapporto all'estensione crescente delle sue applicazioni e all'ambivalenza dei suoi fini.

Essa può valere come ausilio per la vita, il lavoro e le decisioni umane; come macchina che si sostituisce alla persona; come strumento di elevazione o di servitù; come captazione e stimolo abnorme del desiderio. Le posizioni più radicali sull'IA sostengono che essa sostituirà l'intelligenza umana e infine che trasformerà la natura o essenza umana, offrendo uno sbocco finale alla rivoluzione antropologica in atto dovunque. Tali prospettive difficilmente si possono assimilare al trentennale dibattito sul "postumano" e il "transumano", che solo in lieve misura si riferivano all'IA e che con l'esplosione di quest'ultima sono regrediti quasi di colpo. Nel mio contributo svolgerò questioni di riconosciuta importanza sull'IA, muovendo dall'elemento antropologico-teoretico e toccando poi talune problematiche etiche, politico-giuridiche e sociali, il cui numero aumenta così velocemente che è arduo tenere il passo. Impegnativa è perciò la necessità di oltrepassare ogni quietismo a buon mercato, quell'attitudine dello spettatore passivo che sta a vedere quello che succede, perché al momento i rischi sembrano superare le opportunità.

1. Intelligenza umana e intelligenza artificiale non funzionano in modo analogo

Esiste una *differenza inoltrepassabile* tra la conoscenza dell'IA, operata da macchine senza vita né corpo, e l'intelligenza umana che si dispiega con la mediazione del corpo umano vivente. Non è mai troppo mostrarlo, perché la valanga informatica che ci investe quotidianamente non favorisce una riflessione accurata, anzi la ostacola, per cui il discernimento intellettuale è tanto più necessario quanto più è declinante. Per orientarci, occorre porre in evidenza le fondamentali differenze tra macchina/IA e la persona come vivente umano. La macchina è un *insieme di non*: *non* pensa ma calcola, *non* ama, *non* ha corpo e quindi *non* è un vivente, *non* manifesta sentimenti, *non* possiede la nozione di morte, *non* è in grado di dubitare e *non* distingue il bene dal male. L'algoritmo manipola un'enorme quantità di segni e può con

ciò simulare sentimenti che non prova né potrà mai provare. Conosce i segni e le parole per esprimerli, perché il programmatore ha operato così, e perciò li finge ma non li esperisce. In sostanza una scelta apriorica e senza adeguato discernimento a favore della macchina non è un buon affare, perché ci consegniamo indifesi al declino intellettuale: è declino intellettuale smarrire il significato dell'eccezionalità umana e assimilare la persona a macchina. Di fatto potremmo cedere il passo a coloro che possiedono e programmano l'IA, senza valutare il rischio che questi acquisiscano un potere occulto su di noi.

Naturalmente nell'indagine sulla Tecnica e l'IA bisognerebbe introdurre la capitale questione se la tecnica possa tutto, e la risposta negativa che il filosofo pronuncia: per quanto alta possa essere la volontà di potenza esercitata dalla Tecnica, sussistono obiettivi che sono tecnicamente fattibili, e altri che non lo sono né mai lo saranno. Cambiare attraverso l'intervento tecnico l'essenza o la natura umana è qualcosa che oltrepassa la volontà di potenza degli uomini¹. Dopo i "non" di cui sopra chiederemo ora se l'IA sia in senso proprio un'intelligenza, o se tale termine sia profondamente improprio e falsante. Il punto più fondamentale si individua nello stabilire le operazioni proprie dell'intelletto umano e dell'IA, ossia se essi operano secondo modalità confrontabili, oppure nettamente diverse e secondo un raggio di conoscenza non comparabile. L'intelligenza è la sede del pensiero, per cui la prima questione suona: che cosa significa pensare (il vero)? Pensare è *concepire, percepire e giudicare* (l'ordine o comando, l'invocazione, la preghiera, l'offesa, etc. appartengono al pensiero in modo diverso). Pensare è in misura secondaria riducibile a *calcolare*.

L'intelletto umano, con l'intuitività e l'intenzionalità che gli ineriscono necessariamente, ha come orizzonte la totalità dell'essere, il finito e l'infinito. Il mistero primario della conoscenza umana sta nell'identità

¹ Ho svolto in varie occasioni tale tema basilare su cui stranamente tanti non si pronunciano, e in specie quelli critici verso la onnipotenza della tecnica. Il primo intervento rimonta a trenta anni fa: *Le Società Liberali al bivio*, Marietti, Genova 1991. Cf. anche E. SADIN, *Critica della ragione artificiale. Una difesa dell'umano*, LUISS University Press, Roma 2019.

intenzionale tra "pensiero" ed "essere", per cui *intellectus in actu et intellectum in actu sunt idem* nell'ordine intenzionale del pensare (l'intelletto in atto di conoscere e la *res* conosciuta fanno uno nel momento sorgivo della conoscenza). In questa fase primordiale pensiero ed essere entrano in intimo contatto di modo che la prima attività del pensiero è la generazione del concetto, ossia il concepire. La cosa "A" possiede perciò una doppia esistenza: quella naturale là fuori e quella intellettuale all'interno del pensiero, in cui accade l'unione intenzionale in atto tra conoscente e conosciuto. Jacques Maritain così chiarisce in *Distinguere per unire o I gradi del sapere*:

«L'intenzionalità non è soltanto quella proprietà della mia coscienza di essere una trasparenza indirizzata, di prendere di mira gli oggetti in seno a se stessa: essa è prima di tutto quella proprietà del pensiero, privilegio della sua immaterialità, per mezzo della quale l'essere per sé posto "fuori di esso", cioè pienamente indipendente dall'atto stesso del pensiero, diviene esistente in esso, posto per esso e integrato al suo atto: è insomma quella proprietà per mezzo della quale pensiero ed essere esistono in una sola e identica esistenza soprasoggettiva»².

Componendo i concetti, l'intelletto pronuncia un giudizio ("S è P") che sarà vero o falso secondo che si conforma o meno alla realtà, e perciò nel giudizio accade una percezione di conformità o difformità tra intelletto e *res*. Che ne è dell'IA circa il concepire e il percepire? Essa non è in grado né di esercitare un'intuizione intellettuale, né di astrarre l'intelligibile dal sensibile, né di operare l'identità intenzionale tra pensiero ed essere. Essa dunque non concepisce, non sperimenta i *qualia*, non ha come orizzonte il tutto; è una macchina *calcolante*

² J. Maritain, *I gradi del sapere*, Morcelliana, Brescia 2013, 132. La nozione di intenzionalità è lungi dal ridursi al riferimento obbligato a un oggetto, nel senso ovvio che il sapere è sempre sapere di qualcosa, e così il pensiero e ogni altro atto intenzionale che non può non avere oggetto. Questo modo diffuso di presentare l'intenzionalità è a un tempo vero e seriamente riduttivo, perché non include il punto più decisivo consistente nell'intima unione fra conoscente e conosciuto, spesso tralasciato, che conduce alla loro *identità intenzionale* nel concetto. Il significato largo di intenzionalità come "direzionalità mirata a" sta alla base della differenza tra atti o stati mentali e atti o stati fisici, dal momento che i primi sono intenzionalmente diretti verso qualcosa, a differenza dei secondi.

di grandi capacità e in certo modo generativa nel senso che, avendo accesso a un'immensa quantità di dati, può comporli in vario modo, dando l'impressione di creare qualcosa. In realtà l'atto del pensare accade nella sfera dello *spirito come prassi immanente*, capace di ospitare in sé il tutto (*anima est quodammodo omnia*) e ciò non è alla portata del macchinico³. Il funzionalismo non sembra in grado di rispondere alla domanda: "che cosa significa pensare?" se non assimilando brutalmente il pensiero a calcolo⁴.

L'analogia tra intelletto e computer calcolante appare perciò altamente problematica; non basta aggiungere l'aggettivo artificiale per distinguere le due realtà. L'IA non è un'intelligenza, neppure quando, per salvarsi la coscienza, si aggiunge "artificiale". L'intelligenza sgorga dalla vita. La macchina calcola le corrispondenze statistiche ma non sa ciò che fa. E il "realismo dell'intenzionalità" non ha nulla a che vedere col "realismo materialistico" che assume la mente come mero specchio del dato e della natura. La mente non sta al cervello come il *software* all'*hardware*. L'equiparazione funzionalistica fra *software* e mente trascura che la mente è anch'essa vita: vita dell'intelletto, intenzionalità, conoscenza, mentre il *software* è programma, algoritmo formale, modello di procedimento.

Il termine stesso d'intelligenza artificiale è pertanto fuorviante, è un *ossimoro* in cui l'aggettivo contraddice il sostantivo perché lascia intendere che l'intelligenza umana possa essere approssimata dalla macchina. Ciò è vero solo per le attività computazionali della mente, ma fortemente riduttivo per l'impressionante varietà e ricchezza delle espressioni dell'intelletto umano: conoscitive, morali, autocoscienti, estetiche, spirituali, mistiche, musicali, ecc. Né si può omettere che all'intenzionalità conoscitiva si affianca quella dell'azione con le prerogative che le competono: la capacità di autodestinarsi, di stabilire a sé stesso dei fini

³ Maggiori sviluppi sono nel mio volume *Incontrare l'esistenza. Jacques Maritain e la metafisica*, Studium, Roma 2023.

⁴ ChatGPT: strumento di elaborazione del linguaggio naturale e capace di apprendimento automatico per generare risposte simili a quelle umane entro un discorso.

e di indirizzare ed esercitare la libertà. In breve l'intelligenza è vita e non macchina; e se è macchina non è intelligenza. Assumere la prospettiva impersonale sottesa alla prassi scientifica comporta un prezzo pesante per la persona, coinvolta in un esteso processo di naturalizzazione. L'unicità della persona non può in alcun modo essere riportata a un insieme di dati, per quanto sterminato esso sia. Ben poco di ciò che è essenzialmente umano può essere calcolato e misurato:

«Il rispetto fondamentale per la dignità umana postula di rifiutare che l'unicità della persona venga identificata con un insieme di dati. Non si deve permettere agli algoritmi di determinare il modo in cui intendiamo i diritti umani, di mettere da parte i valori essenziali della compassione, della misericordia e del perdono, o di eliminare la possibilità che un individuo cambi e si lasci alle spalle il passato»⁵.

2. L'IA e la "persona"

Dedichiamo ulteriore attenzione al confronto tra IA e intelligenza *naturale umana*, propria della persona. Quando si allude alla seconda si pensa forse di conoscerla adeguatamente: illusione perniciosa perché la conoscenza approfondita dell'*intellectus* e della *ratio* della persona (due operazioni della stessa facoltà cognitiva) è molto ardua e non saranno le sole misurazioni empiriche a risolvere il problema. L'intelligenza umana è una facoltà che, pur assumendo dalle sensazioni il contenuto o informazione intelligibile che esse veicolano, possiede una vita superiore e ulteriore a quella della materia animata. È una facoltà spirituale. Supponiamo che vi sia vita su Venere, ma ciò non implica affatto che con ciò vi sia in essa intelligenza. L'*intelligere* è una vita, certo, ma infinitamente superiore alla mera vita probiotica elementare, perché in esso si conosce l'essenza e l'esistenza delle cose (enti). Vi sono certo altre forme di intelligenza pratica, pragmatica, affettiva, emotiva che possiamo in una qualche misura avere in comuni con i primati, ma non di più.

⁵ PAPA FRANCESCO, *Messaggio per la 57ma Giornata mondiale della pace*, 1° gennaio 2024.

La forma propria della vita umana è di vivere secondo ragione o *logos*.

Le macchine né vivono né muoiono; funzionano per un certo tempo, e poi si rompono o sono superate da altre macchine di nuova generazione. Inoltre l'IA non ha in se stessa intenzioni e fini, che vi sono invece inseriti dall'essere umano; per questo non dobbiamo diffidare della macchina ma di noi stessi. L'IA non replica (produrre una copia identica all'originale) né simula (produrre un modello che si accosti molto all'originale) la mente umana, e neppure la emula (cercare di eguagliare un comportamento). Invece di rivolgersi all'IA chiedendole "miracoli" che non potrà compiere, risulta più saggio interrogare noi stessi su come intendiamo impiegarla. Si dice che l'IA prenderà ogni potere, ma bisogna aggiungere che, se sarà così, la responsabilità sarà nostra. Non dobbiamo preoccuparci delle varie IA, ma soprattutto di coloro che le progettano, perché non conosciamo i loro scopi reali se non a posteriori.

3. Sull'"aumento della potenza"

Da molte parti sorge la richiesta ragionevole di sviluppare un'IA etica, che non superi i limiti stabiliti pubblicamente dalle diverse comunità politiche. A mio parere ancor prima di entrare nel *mare magnum* di tali problematiche occorre meditare su una domanda etico-antropologica decisiva: l'aumento della potenza è un bene in sé? Molto ci distoglie dall'offrire una risposta positiva all'interrogativo. L'incremento della potenza è intrinsecamente ambivalente poiché può essere impiegato per migliorare la condizione umana tanto quanto per asservire e umiliare: *De omni technica dubitandum est*, perché essa è ambivalente. Dubitare significa fermarsi a riflettere e considerare ciò che siamo: se stiamo seguendo un cammino dotato di senso o meno, se siamo trascinati dagli eventi o meno, se stiamo esercitando un pensiero critico. Bisogna pertanto mettere in atto, senza paura di essere considerati conservatori privi di coraggio, una diffidenza intelligente verso il vasto mondo dell'IA. L'utente dei *media* e dell'IA deve interrogarsi sugli scopi che stanno a

cuore al "produttore". Questo si articola in una catena complessa in cui incontriamo l'imprenditore che investe i suoi capitali, il programmatore che deve cercare il programma che vada incontro ai gusti predominanti e/o ne crei di nuovi, la rete che diffonde, il pubblicitario, la concorrenza per la conquista di quote di mercato, etc.

Collocandoci a un livello più profondo rimane basilare oggi come ieri, ma si potrebbe anche sostenere oggi più di ieri, la presenza ispirante, regolante e normativa di un'idea adeguata della persona umana che il personalismo ontologico/metafisico può offrire. La conclamata fine della metafisica e il trionfo della tecnica hanno tra loro molti legami causali. Esaminata attentamente, l'IA invece di svalutare la condizione umana e l'eccezionalità dell'esser-persona, le conferma ad ogni passo. Come non vi è bisogno di abbassare l'uomo a livello animale per elevare l'animale, altrettanto assurdo è abbassare l'uomo a macchina povera e inefficiente per mettere sull'altare l'IA.

L'IA lasciata a sé stessa produce una nuova forma della gabbia di acciaio della razionalità strumentale, preconizzata da Max Weber, che intuì il problema ma non fu in grado di indicare un'etica del limite antropologicamente ed ecologicamente fondata. La questione fu ripresa da Romano Guardini, che nel 1950 invitava l'uomo contemporaneo a fare i conti e a limitare la potenza: «L'uomo moderno non è stato educato al retto uso della potenza»⁶. Il mito del progresso lineare era già morto ai primi del '900, e sempre di più si è imposta l'idea che la tecnica porta vita e morte, benessere e infelicità. Alla domanda se l'aumento della potenza sia un *bene in sé* è giustificato rispondere negativamente⁷.

⁶ R. GUARDINI, *La fine dell'epoca moderna*, Morcelliana, Brescia 1954, 86.

⁷ Osserva acutamente Giuseppe Limone: «Il mondo della vita, impiegando il Logos, trasforma la sua capacità auto-regolatrice in macchina e in sistemi di macchine. Questa macchina, arrivata a questo grado di evoluzione, è separata dalla vita, è funzionale e automatica. In quanto tale, ignora il *tu*, ciò che è vivente e l'interiorità, ossia opera senza discrezionalità, senza responsabilità e senza dialogicità: è tutta e soltanto *proceduralizzata* [...]. La rigorosa forma organizzativa avrà estinto lo spirito per il quale è stata istituita. Anche in questo caso, come negli altri consimili, sono accadute tre cancellazioni essenziali, tutte invisibili: la rimozione dell'altro, la rimozione del "qui e ora", la rimozione dell'interiorità. Un tale

4. La coscienza

Una parte considerevole della filosofia moderna ha assegnato un ruolo preminente alla coscienza (e autocoscienza), facendone in certo modo l'essenza dell'esser-persona, come nell'idealismo tedesco e nell'attualismo italiano. Quando questi sono retrocessi, è rimasta forte la visione funzionalistica della persona: questa c'è solo se coscienza e autocoscienza sono presenti⁸. È noto che da numerosi decenni si sono messe in opera varie strategie per ridurre la coscienza e/o la mente al cervello⁹. E poi per sviluppare reti neurali che simulino le operazioni della mente-cervello.

Toccando il tema della coscienza occorre ricordare che esso è polivo-co (dai molti significati), possedendone almeno due: la coscienza *psicologica* (in cui potremmo includere anche l'autocoscienza) e la coscienza *morale*. Certamente la seconda è estranea alla macchina, ma anche la prima deve essere revocata in dubbio. Gli animali la possiedono in varia misura secondo le specie, ma non le macchine che non hanno vita, e la coscienza appare con la vita, non mediante segnali elettrici. La coscienza include l'autocoscienza, ossia la capacità di entrare nella propria interiorità e di autocomprendersi, scendendo sin che si può (in genere non si può molto) *nel fondo senza fondo di sé stessi*. L'io umano non è in grado di scendere in sé stesso per toccare la sorgente dell'anima, quel luogo inaccessibile da cui si dipartono le facoltà della persona. In tutto ciò gioca nei viventi e in specie nell'uomo la corporeità e il suo immenso rilievo per la percezione e l'autopercezione. È impossibile supporre un funzionamento dell'intelletto *umano* senza il *corpo*. I naturalisti e i

risultato non è un incidente di percorso, ma una deriva strutturale» (G. LIMONE, *L'algoritmo e il mondo della vita. Nuovi appunti sul fondamentalismo macchinico nell'era contemporanea*, in *Persona. Periodico internazionale di studi e dibattito*, n. 1-2 (2017) 36-37.

⁸ Da questo postulato alcuni autori hanno dedotto che l'infanticidio poco dopo la nascita sarebbe lecito, poiché non sarebbe presente la persona. In tal modo verrebbe riformulato il quinto comandamento – non uccidere – diventando "non uccidere la persona".

⁹ Su questi nuclei rinvio ai miei volumi *Il Nuovo Principio Persona*, Armando, Roma 2013 e *La rivoluzione biopolitica. La fatale alleanza tra materialismo e tecnica*, Lindau, Torino 2013.

materialisti ignorano l'importanza del primo, gli intellettualisti-razionalisti si fanno beffe del corpo, e in entrambi i casi l'esito sarà un rischio in più per la persona nel suo rapporto con la tecnica. La persona umana in quanto dotata di un intelletto incorporato include funzioni che l'IA non potrà compiere. Insomma per valutare l'IA non è sufficiente richiamare la grande utilità di impiegarla per edizioni critiche di libri, raccolta dati, impiego nel turismo, nell'industria, etc.

5. L'IA e la vita di tutti i giorni

Il contatto quotidiano con il mondo digitale tende a offuscare la diversità tra virtuale e reale, operando una trasformazione ambigua dell'esperienza umana e del senso comune. Talvolta si crede che in tanti dilemmi decida meglio l'IA invece che l'uomo. Bisogna perciò evocare alcuni punti nodali, segnalati fortunatamente da molti, sulle controindicazioni dell'IA nel suo rapporto con la persona. Ne trascelgo tre: il declino (sino alla perdita) del pensiero critico, che consegue non dall'imitazione meccanica della mente umana bensì dall'appiattimento di questa sulla macchina. La mente rischia di formarsi non nel confronto col reale ma con l'IA, la quale è un pensiero binario (sì/no) che perde le tante sfumature e la complessità della realtà. In certo modo il soggetto si vergogna di essere lento dinanzi all'iper velocità della macchina, e non dovrebbe farlo perché la ricerca e il pensiero lenti sono necessari per approfondire e valutare. Il secondo equivoco sta nell'assunto che conoscere significhi soprattutto accumulare dati e conoscenze puntuali, in un processo additivo indefinito. Infine all'impoverimento del pensiero discernente si aggiunge un deficit di creatività, perché l'algoritmo ci conduce verso cammini prefissati, indebolendo la naturale curiosità dell'essere umano, e il valore dell'imparare con sacrificio. Ci priviamo della gioia che proviene dall'aver compiuto un cammino nostro con successo. Non dobbiamo sottovalutare l'avvento della pigrizia: avizzeremo domande e attenderemo pigramente le risposte dalle tecnologie. A titolo cursorio aggiungo altre controindicazioni. La disincarnazione delle relazioni

umane probabilmente aumenterà con il ricorso crescente all'informatica. La nozione di luogo sarà esposta al declino: già adesso siamo "altrove" rispetto al "qui" e "ora" in cui esistiamo. Non consistere più in un "dove" è una modalità dello sradicamento e della depersonalizzazione, che favorisce la solitudine delle persone carenti di rapporti reali. L'abilità manuale, già ridotta da tempo nei giovani e negli adolescenti, potrà diminuire e condurre a un minor senso della componente corporea e materiale dell'esistenza, e forse a un declino dell'interazione tra manuale e mentale. Di pari passo potrebbe procedere il declino della capacità di autodeterminarsi e di nutrire un pensiero meditato, non distorto dalle decisioni immediate suggerite dagli algoritmi. In questo ambito di immanenza inclusiva rischiamo di perdere il sensorio della Trascendenza e dell'alterità.

Con i LLM (*Large Language Models*) l'IA acquisisce capacità generative nel comprendere ed elaborare il nostro linguaggio di esseri umani nel *deep learning*. I risultati possono essere sorprendenti, e in ogni caso è necessario valutarli, sulla scorta di un pensiero discernente che non si acquista agevolmente. Bisognerebbe continuare ad acquisire conoscenze attraverso la lettura di libri stampati e l'ascolto delle lezioni degli insegnanti. Questa avvertenza vale in special modo dinanzi allo sviluppo cognitivo del bambino e dell'adolescente, che richiede la formazione del ragionamento logico, la lettura distesa e calma, la riflessione e l'esercizio della memoria. In caso contrario le capacità del bambino e dell'adolescente vengono indebolite e i rischi di analfabetismo linguistico e culturale crescono. Da decenni molti osservatori hanno rilevato la stupefacente povertà del linguaggio di tanti, la difficoltà di articolare il pensiero in maniera adeguata, di comprendere testi e brani di media difficoltà. Non si impara a scrivere e a esprimersi in maniera accettabile se non scrivendo e parlando in prima persona. Su questi aspetti si gioca la sorte di un popolo, perché non dura a lungo come popolo un insieme di soggetti che non sono più capaci di esprimere i propri sentimenti e passioni in maniera articolata, non rozza e segnata dalla violenza verbale.

6. Tecnica, etica e diritto

Tutti constatiamo che sussiste una costante precedenza di fatto della tecnica rispetto all'etica, che come l'intendenza viene dopo. Essa è alla rincorsa senza riposo delle nuove possibilità tecniche che crescono ogni giorno di più e che rendono visibile la debolezza prescrittiva della morale. Lo sappiamo da più di un secolo, da quando Henri Bergson chiedeva un supplemento d'anima per essere in pari con lo sviluppo tecnico. La difficoltà a regolare la tecnica non è da allora diminuita ma aumentata. Volenti o nolenti ci adeguiamo al processo continuo, cercando di trovarvi un senso e un orientamento che perlopiù sfuggono. Non sappiamo dove dirigerci poiché lo spirito prometeico che ci abita non dà respiro. Né desideriamo come Ulisse tornare ad Itaca: le radici hanno perso valore ai nostri occhi trascurati. Una tensione quasi lacerante abita le nostre vite, in specie in occidente, divise tra un intento umanistico e la volontà di potenza che alimentiamo in noi¹⁰.

Nell'ambito del diritto continua a essere inalienabile l'assunto che la persona umana debba essere giudicata da un'altra persona, non da una macchina o un'IA gestita da Stati e Corti. Qui la pervasività del mondo digitale va attentamente controllata perché, se lasciata a sé stessa, opera contro il principio cui ci siamo appellati. La decisione algoritmica non può essere esclusiva, ma deve prevedere il controllo umano, che potrà convalidare o smentire la decisione della macchina. Al momento una Commissione dell'UE ha emanato una carta etica sull'uso dell'IA nei sistemi giudiziari per normare la giustizia predittiva e la *Prediction Technology* che calcolano la probabilità dell'esito di una controversia giuridica.

¹⁰ Sulla crisi dell'occidente si veda il mio contributo *Individualismo, libertismo, stanchezza della storia in Occidente e in Europa*, in *I nodi dell'Occidente*, a cura di M. De Angelis, Belforte, Livorno 2023, 131-148.

7. Nessun quietismo né arrendevolezza

L'assunto che l'IA e l'intelligenza umana differiscono senza possibilità di identificazione non deve indurre al quietismo. Rischi e opportunità si aprono quasi contemporaneamente quando si cerca di restaurare e potenziare l'essere umano, sia curando malattie sia dotandolo di maggiori capacità. In questo campo possono accadere eventi buoni o cattivi. Consideriamo il cambiamento iperveloce del tessuto esistenziale e le difficoltà di molti di reggerne il ritmo, con le conseguenti fratture sociali in molti campi. Senza un'idea adeguata della persona, dei suoi diritti e doveri, della sua dignità, la volontà di potenza della tecnica – che in realtà è volontà di potenza dei singoli e dei grandi gruppi e *holdings* che operano poderosamente su scala mondiale, spesso in un serio vuoto normativo – genera violenti squilibri. Finora scarsa è stata la capacità dell'autorità pubblica di regolamentare i grandi produttori di IA che, costituiti da poderosi gruppi privati egemoni a livello mondiale, mostrano un'alta riluttanza a sottoporsi a controlli e normative.

Un esempio del cattivo impiego dei *media* balza agli occhi se si riflette sull'odio che circola sulla rete e che cresce autosostenendosi, e in cui il paradigma dominante è l'opposizione tra amico e nemico. L'espressione dell'odio, intensificando il ricorso al mezzo, rende economicamente molto più di altri business, senza che si calcolino i danni inflitti in specie ai minorenni e ai bambini che crescono in tale clima. Una volta di più si mostra vero che i rischi per l'umanità non vengono da errori delle tecnologie, ma dal loro uso malsano. Ogni tecnica è aperta sui contrari, sul suo uso buono o cattivo, e ciò non dipende dalle tecnologie ma dall'uomo che le progetta e le impiega. L'energia atomica illumina le città ma può essere impiegata per distruggerle. Il *chip* che viene installato nel cervello non solo consente di interpretare i segnali elettrici di coloro che non possono comunicare con l'esterno, fornendo un aiuto ma consente parimenti di inviare segnali esterni al cervello, con il rischio di manipolazione e di espropriazione del soggetto. Non si dovrebbe mai dimenticare l'intrinseca *ambivalenza* della

tecnica. Rientra in questo discorso la responsabilità morale delle aziende che diffondono notizie false, magari per incrementare i loro profitti o per influenzare negativamente l'opinione pubblica, e l'opportunità che siano sanzionate.

Siamo perciò rinviiati alla questione del potere dei *media* e degli *imperi digitali*. La loro grande capacità di comunicazione ad alta velocità e le agevolazioni che se ne traggono non devono nascondere agli occhi di tanti che dietro le piattaforme più diffuse vi sono capitali immensi e potentati economici che hanno i loro scopi: il massimo profitto, la omogeneizzazione del pensiero che può giungere sino al controllo. Chiediamo: a chi spetta l'ultima parola sul possesso e l'esercizio dell'IA? La soluzione che si profila in occidente è un acuto oligopolio dell'IA, in mano a pochissime aziende con capitali giganteschi. Nelle democrazie tale irrisolta questione comporta un sovraccarico informativo e algoritmi preposti a selezionare le informazioni che si decide di far circolare. Nasce un nuovo *tecnofeudalesimo*, come lo chiama Yanis Varoufakis¹¹ (*La nave di Teseo*, 2023), capace di controllare e modificare i comportamenti delle persone, infeudandole a pratiche che consentono ai proprietari profitti enormi. Altro grande rischio sta nella creazione di programmi e messaggi digitali che in modo subliminale spingano verso l'uso compulsivo ed esteso di varie piattaforme.

E infine la domanda centrale: se l'IA provoca danni, chi ne risponde? Il progettista, il venditore, l'utente? Le grandi piattaforme e i loro proprietari devono riconoscere e accettare quello che finora hanno tenacemente negato, ossia la responsabilità sui contenuti che trasmettono – essi sono editori e non solo canali tecnici di trasmissione – che influiscono in misura crescente su bambini e adolescenti favorendo l'insorgere di sintomi di autolesionismo, di depressione, di perdita di autostima, e inclinazione verso la violenza e l'odio sociale¹². Accurato esame

¹¹ Y. VAROUFAKIS, *Tecnofeudalesimo. Cosa ha ucciso il capitalismo*, La nave di Teseo, Milano 2023.

¹² Il presidente Sergio Mattarella ha ricordato varie volte tali rischi, e da ultimo nel suo messaggio di fine anno (31 dicembre 2023): «Penso anche alla violenza verbale e alle

richiedono poi i sistemi di IA in ambiti ad alto rischio quali i settori della salute, della sicurezza e dei trasporti.

I *Cloud Empires*, ossia le grandi piattaforme che possono per alcuni aspetti assomigliare agli Stati, se ne differenziano però profondamente perché non debbono rendere conto del loro operato ai cittadini: *non vi è o non sussiste una democrazia digitale*. Non vi sono al momento capacità adeguate di controllo delle grandi piattaforme multinazionali, specialmente in occidente. In Cina le imprese di IA sono soggette al controllo pervasivo del Partito unico; negli Usa dove vige la libertà di mercato, il loro potere risulta oggi soverchiante e poco soggetto a controlli e regolazione.

Un bisogno forte che avanza starebbe in Carte costituzionali digitali dei diritti e doveri fondamentali sia delle piattaforme sia degli utenti. Ci dovrebbero essere regole scritte e delle sedi capaci di sanzionare chi le viola. Considerata la globalizzazione dei *media* e dei loro prodotti informativi, sarebbe necessaria una regolazione universale che tocchi tutti i Paesi con normative conformi: una questione al momento quasi insolubile per l'accentuato disordine globale¹³.

espressioni di denigrazione e di odio che si presentano sovente nella rete». Fa riflettere che le espressioni più usate per esprimere tali atteggiamenti siano in larghissima parte di provenienza anglosassone e più esattamente americana. Parole per comunicare disprezzo, derisione, ostilità, insulti, volontà di odio verso l'altro. Il citato documento di papa Francesco evoca i gravi rischi di manipolazione e di controllo sociale, e chiede «una chiara responsabilità legale da parte dei produttori, di chi le impiega e delle autorità governative» (intendo: responsabilità *dinanzi* alle autorità governative). Il messaggio del papa, dopo aver osservato che le intelligenze artificiali imitano o riproducono solo alcune funzioni dell'intelligenza umana, aggiunge che esse costituiscono sistemi sociotecnici che in tanta parte dipendono dagli obiettivi e dagli interessi di chi li possiede e li sviluppa.

¹³ Qualcosa si muove in sede ONU: il segretario generale António Guterres ha creato recentemente un *Advisory Body* dell'Onu sull'IA, per favorire la creazione in sede ONU di un'agenzia deputata a monitorare giorno per giorno, sorvegliare e regolamentare la realizzazione e applicazione dei sistemi di IA, avendo scopo e ampiezza mondiali. Il messaggio di papa Francesco invita la comunità delle nazioni «a lavorare unita al fine di adottare un trattato internazionale vincolante, che regoli lo sviluppo e l'uso dell'intelligenza artificiale nelle sue molteplici forme», mirante a prevenire le cattive pratiche e a incoraggiare le buone pratiche, tra cui l'auspicio che l'IA «non accresca le troppe disuguaglianze

Si dovrebbe creare un'agenzia mondiale di regolazione e normazione per rendere compatibili le varie regolamentazioni nazionali ed estenderne il valore condiviso su scala mondiale. Una tale agenzia, posta accanto alla FAO e all'OMS, costituirebbe un primo nucleo di una futura (e lontana) autorità politica mondiale, nonostante la loro strutturale debolezza in quanto sono organismi messi in opera da Stati sovrani.

8. Conclusioni

Fermare l'accelerazione dell'IA non è attualmente possibile, ci rimane la possibilità di orientare e regolare il suo cammino. Il pericolo *sociopolitico* più serio risiede nel fatto che la potenza e gli sviluppi dell'IA sono in Occidente nelle mani di pochissimi gruppi privati dotati di un potere straordinario. Da qui la domanda: può una decente democrazia convivere con una disparità di poteri come quello che si preannuncia, e che di fatto esiste già? La recente vicenda di *OpenAI* sembra si sia conclusa con la vittoria del desiderio di potere e guadagno sugli intenti originari di tenere conto dei dilemmi etici intrinseci all'IA e di tenerla come sorgente "aperta". L'IA diffonde un nuovo ambiente culturale in cui diventa imprescindibile sapersi orientare e muoversi, il che richiede sin dai primi anni di vita dell'essere umano un'educazione sobria e intelligente al suo impiego. Buone notizie vengono dalla Svezia, dove è stato deciso di tornare nella scuola primaria a penne, quaderni e libri, mettendo da parte computer, tablet e lavagne luminose. Il motivo è che gli strumenti digitali abbassano la capacità di apprendimento e di lettura. Ciò avrà conseguenze positive per molti, tra cui i bambini che stanno perdendo la manualità necessaria e la capacità di distinguere tra mondo del computer e mondo effettivo. Il robot non è uno schiavo poiché gli schiavi erano e sono persone, è una macchina progettata

e ingiustizie già presenti nel mondo».

dall'uomo per essere utile all'uomo¹⁴. E in quanto tale deve acuire il senso e la qualità della responsabilità della persona verso gli altri, se stessi, le generazioni future e l'ecologia. Il principio di responsabilità di Hans Jonas va pertanto tenuto presente e attivo in ogni considerazione etica concernente l'impiego delle tecnologie. Sarebbe un passo falso ritenere che il progresso tecnologico sgravi le coscienze dal peso della responsabilità; anzi tale progresso esige dalle persone un supplemento di responsabilità e di opzioni morali, per le quali la sola etica di impianto utilitaristico e del "miglior interesse" non è sufficiente.

La sfera dell'IA è la sfera dell'umano, delle nostre domande, tradizioni, forme di vita, e l'IA si muove dentro questa sfera, apportando miglioramenti e peggioramenti; ossia non crea un nuovo universo morale, teoretico e immaginativo diverso dalle forme fondamentali dell'esser-persona. Non ci sarà un successore dell'uomo nel dominio della Terra, e non dobbiamo temere che entro un certo periodo l'*homo sapiens sapiens* sarà declassato a razza inferiore. Vi è timore dinanzi all'IA, ma non perché le macchine diventeranno autonome e prenderanno il sopravvento declassando l'uomo – una fantasticheria senza fondamento, disapprovata dalla scienza e dalla filosofia – quanto perché gli esseri umani potranno usarle male, infliggendosi danni e disgrazie.

Un indebolimento del pensiero critico e della capacità di giudicare potrà provenire dalla disinformazione e dalle notizie false spacciate per vere. In tale quadro possono rientrare i "ladri di attenzione" che mirano a catturarci senza requie per i loro scopi, spesso finalizzati a intenti di manipolazione e di guadagno, o anche a ristrutturare permanentemente i nostri desideri. Non dobbiamo farci rinchiudere nelle profilazioni costruite su di noi e a nostra insaputa dai grandi gestori delle reti informatiche. Una cultura sanamente critica da parte di intellettuali non schierati pregiudizialmente rappresenta una necessità, capace anche di individuare e contrastare la creazione mediatica di falsi bisogni.

¹⁴ Vedi R. PEZZIMENTI, *De veritate. Cinque prospettive di ricerca*, Armando, Roma 2023, in specie il cap. V.

Va infine rilevato come l'ombra lunga della "postverità" comporti la trasformazione della nozione di "documento". Questo come fonte affidabile riposa su due elementi: l'autenticità del testo (collocazione temporale, supporto come la carta o la pergamena, attribuzione di paternità) e la correttezza dei contenuti in rapporto al contesto in cui il documento è stato elaborato, per cui noi non attribuiremmo correttezza di contenuti a un eventuale testo scritto in greco antico, o in etrusco, che parlasse di computer. Osserva bene Vincenzo Ambriola (*Il potere della parola messo in dubbio dall'IA*, in *Avvenire*, 23 dicembre 2023, 18) che l'intelligenza artificiale generativa (ChatGPT) ha messo in discussione il concetto stesso di "documento": «La novità, inattesa e sconvolgente, è la capacità dell'intelligenza artificiale di *generare* contenuti simili a quelli "umani" in maniera stocastica (ovvero non deterministica e su base probabilistica) a partire da dati di addestramento spesso sconosciuti». Pertanto la generazione artificiale di documenti che non hanno una vera paternità e di cui si può mettere in dubbio l'autenticità apre il vaso di Pandora della "notizie false" che pullulano nelle reti social, nel *web* e nei *media*. Con le tecniche attuali quasi tutto può essere falsificato e risulta di crescente difficoltà separare il vero dal falso, con la conseguenza di trovarci indifesi dinanzi all'errore e alla falsificazione. Il problema concerne in modo immediato i docenti e gli allievi nel loro cammino: «Si deve affermare con chiarezza e semplicità che l'intelligenza artificiale non è una fonte affidabile di sapere e di conoscenza, ma che è soggetta ad allucinazioni che le fanno anche generare contenuti errati o, a volte, totalmente inventati».

DOMAIN SPECIFIC MACHINE LEARNING NELLE SCIENZE UMANE

Riccardo Pozzo

Abstract

L'IA non dà informazioni su tutto, ma risponde a precise domande poste da ricercatori che lavorano su domini specifici. La nuova edizione dell'Accademia delle opere di I. Kant offre qui l'occasione per riflettere sull'IA nella storia della filosofia. A una prima parte sull'interazione uomo-macchina seguono proposte metodologiche e quindi la presentazione della nuova edizione di Kant. Si conclude con osservazioni sul ruolo che la storia della filosofia ha per l'innovazione e la co-creazione.

Parole chiave: Co-creazione, Interazione uomo-macchina, Kant, Storia della filosofia, Apprendimento automatico.

AI does not give information about everything, but answers precise questions posed by scientists working on specific domains. The Academy's new edition of I. Kant's works is an opportunity to reflect on AI in the history of philosophy. A first part on human-machine interaction is followed by methodological proposals and by the presentation of Kant's new edition. The fourth part concludes with remarks on the role that the history of philosophy has for innovation and co-creation.

Keywords: Co-creation, Human-machine interaction, Kant, History of philosophy, Machine learning.

1. Interazione uomo-macchina

L'intelligenza artificiale non dà risposte a tutto. Le risposte che dà sono in realtà il risultato di domande poste da scienziati che operano su determinati domini specifici e che hanno costruito dei *corpora* di

dati e li hanno addestrati (*domain specific machine learning*). Abbiamo ancora ben presente la riconoscenza che si deve a quei gruppi di scienziati che durante la pandemia Covid-19, dopo aver costruito i rispettivi *digital twins*, hanno addestrato i dati che simulavano le proteine fino a piegarle (*protein folding*) per farle interagire con una simulazione del virus, permettendo così di individuare quelle con il principio attivo più efficace per la preparazione del vaccino; e con ciò, è bene notarlo, la *computational biology* ha fatto risparmiare mesi se non anni di ricerche condotte secondo i metodi tradizionali¹.

Nel dominio specifico delle scienze umane, l'intelligenza artificiale lavora su *corpora* preparati come *Large Language Models*. In questo caso, l'obiettivo non è piegare delle proteine, ovviamente, ma piuttosto integrare e talvolta sfidare la ricerca qualitativa basata sulla lettura ravvicinata (*close reading*), per confermare o rifiutare quantitativamente ipotesi qualitative oppure addirittura iniziare la ricerca partendo da risultati quantitativi (*distant reading*)².

2. Metodologia

Tradizionalmente, la filosofia si articola in tre discipline principali: la *filosofia teoretica*, la *filosofia morale* e la *storia della filosofia*³. Delle tre, la storia della filosofia e la sua più recente cugina, la filosofia "comparativa"⁴, sono in questi ultimi anni diventate oggetto di severe critiche e sono cadute in notevole discredito. Anche se detto discredito è dovuto

¹ <https://alphafold.ebi.ac.uk> (accesso 7.12.2023). Cf. J. SYMONS, *What is Domain Specific Knowledge? What Limits to Machine Learning Teach us about Inquiry*, in M. BAGHRAMIAM (éd), *Connaissance et réalité/ Knowledge and Reality. Entretiens de Dublin de l'Institut International de Philosophie* (2023), in preparazione.

² F. MORETTI, *Distant Reading*, Verso, London 2013.

³ <https://plato.stanford.edu/> (accesso 7.12.2023); P. STEKELER-WEITHOFER, *Philosophiegeschichte*, de Gruyter, Berlin/Boston 2006.

⁴ S. BURIK - R. SMID - R. WEBER (eds), *Comparative Philosophy. Contemporary Practices and Future Possibilities*, Bloomsbury, London 2020.

a diverse ragioni⁵, una spicca sulle altre: che la ricerca *mainstream* nella storia della filosofia e nella filosofia comparativa è stata dominata da una metodologia basata su diverse forme di *close reading* che in questi ultimi anni sta mettendo capo a risultati che appaiono deficitari non solo rispetto all'evidenza empirica e al *data-driven* delle scienze ma anche rispetto al rigore delle altre parti della filosofia⁶.

Oggi l'intelligenza artificiale permette un radicale cambiamento di prospettiva, poiché l'aver configurato molti testi filosofici (letteratura primaria e secondaria) nella forma di dati lemmatizzati rende possibile aprire una dimensione quantitativa ed empirica anche nella storia della filosofia e nella filosofia comparativa. È interessante notare che a partire dall'*Index Thomisticus* di Roberto Busa⁷ diversi gruppi di storici della filosofia sono stati pionieri utilizzando inizialmente uno scrupoloso approccio manuale⁸, applicando poi i primi metodi computazionali⁹, e finalmente esplorando l'interazione uomo-macchina nella considerazione dei testi filosofici¹⁰.

Oggi possiamo utilizzare le tecnologie di apprendimento di funzioni e rappresentazioni *self-supervised* per "lasciare che i dati parlino da soli". È dunque utile pensare a un nuovo approccio empirico, che riprende la scelta programmatica del *Lessico Intellettuale Europeo* di rispettare

⁵ R. RORTY, *The Historiography of Philosophy. Four Genres*, in R. RORTY - J. B. SCHNEEWIND - Q. SKINNER (ed.), *Philosophy in History*, Cambridge, Cambridge University Press 1984, 49–76. <https://doi.org/10.1017/CBO9780511625534.006>; T. SORELL - G. A. J. ROGERS (ed.), *Analytic Philosophy and History of Philosophy*, Clarendon Press, Oxford 2005.

⁶ C. MERCER, *The Contextualist Revolution in Early Modern Philosophy*, in *Journal of the History of Philosophy* 57 (2019) 529-548. DOI: 10.1353/hph.2019.0057 [open access]; R. POZZO, *History of Philosophy and the Reflective Society*, de Gruyter, Berlin/Boston 2021. DOI: 10.1515/9783110709292 [open access].

⁷ R. BUSA, *Index Thomisticus*, Corpus Thomisticum, Milano 1970.

⁸ A. O. LOVEJOY, *The Great Chain of Being*, Harvard University Press, Cambridge (Mass.) 1936.

⁹ BUSA, *Index Thomisticus*, cit.

¹⁰ P. VAN WIERST - S. VRIJENHOEK - S. SCHLOBACH - A. BETTI, *Phil@Scale. Computational Methods within Philosophy*, in *CEUR Workshop Proceedings* (2016).

l'esortazione di Montaigne «qu'on définisse rien!»¹¹. Gli utenti identificano dapprima i livelli e le categorie di analisi per poi eseguire annotazioni su un sottocampione del corpus che funge da dato di riferimento e diventa a sua volta la base per ulteriori esperimenti di apprendimento automatico con classi di modelli, includendo strumenti o risorse di dati aggiuntivi¹². L'idea di base è quella già praticata dalla *linguistica dei corpora*, secondo la quale gli allineamenti paralleli multilingue di *corpora* umanistici ricchi di termini tecnici costituiscono materiale assai utile per accelerare l'apprendimento automatico e quindi migliorare l'efficacia della modellizzazione computazionale dei concetti e della traduzione del testo letterario. Oggi gli studiosi possono accedere sinotticamente alle diverse traduzioni da una lingua all'altra e alle diverse definizioni che nel tempo sono state date dei termini.

Considerando le attuali tendenze verso una storia della filosofia *data-driven* come branca delle *digital humanities*¹³, l'idea è che il futuro della storia della filosofia dipenda dalla ricerca di modi per migliorare radicalmente il modo in cui curiamo, archiviamo, annotiamo, accediamo e traduciamo i *corpora*. In questa direzione, il ricorso all'intelligenza artificiale consente anche a chi si occupa di filosofia di confermare o refutare quantitativamente ipotesi qualitative e persino di iniziare la ricerca a partire da risultati quantitativi.

¹¹ M. DE MONTAIGNE, *Les Essais*, a cura di J. BALSAMO - C. MAGNIEN-SIMONIN - M. MAGNIEN, Gallimard, Paris 1963.

¹² J. KUHN, *Computational Text Analysis within the Humanities*, in N. REITER - A. PICHLER - J. KUHN (hrsg.), *Reflektierte algorithmische Textanalyse*, de Gruyter, Berlin/Boston 2021, 61-106, qui 76.

¹³ A. BETTI - H. VAN DEN BERG, *Towards a Computational History of Ideas*, in *CEUR Workshop Proceedings* (2016); A. BETTI - Y. ORTWIN - C. TREIJTEL, *History of Philosophy in Ones and Zeros*, in E. FISCHER - M. CURTIS (ed.), *Methodological Advances in Experimental Philosophy*, Bloomsbury, London 2019, 295-332.

3. La nuova edizione dell'Accademia di Kant

Per fare un esempio concreto, il 31 dicembre 2023 è in uscita il primo volume della nuova edizione dell'Accademia delle opere di Kant¹⁴, che, oltre a offrire la nuova edizione (non uniformata) dei testi da lui scritti tra il 1747 e il 1756, contiene un indice dei nomi, i riferimenti alle citazioni letterali da parte di Kant di autori esterni e i riferimenti incrociati all'interno degli scritti kantiani (ad esempio, dalle *Druckschriften* al *Nachlass* e alle *Vorlesungsnachschriften*). I testi sono stati tutti lemmatizzati in modo molto accurato, con un *markup stand-off* che rende la formattazione del testo, la sua articolazione in livelli, le parole straniere, i nomi propri e i riferimenti incrociati a testi kantiani e non kantiani.

Nel 1894 Dilthey diede vita all'Edizione dell'Accademia degli scritti di Kant per consentire l'accesso alle opere complete di Kant a studiosi e studenti. La terminologia kantiana è stata oggetto di numerose ricerche¹⁵. È diventato presto chiaro che per comprendere e poter giudicare il pensiero di Kant era necessario padroneggiare appieno la terminologia kantiana e la spiegazione più precisa di parole e argomenti¹⁶. I dati relativi a Kant sono stati annotati con crescente regolarità nel corso di oltre cinquanta anni. L'inizio lo diede Gottfried Martin con lo *Allgemeiner Kantindex*¹⁷, che fornisce le concordanze dei testi kantiani in forma non flessa. Un salto di qualità lo permisero Tullio Gregory con il *Lessico Intellettuale Europeo* (che inizialmente utilizzava un markup

¹⁴ I. KANT, *Kant's gesammelte Schriften. Abtheilung I. Werke Neuedition*, Vol. 1: *Schriften 1747-1756*, (hrsg.) H. HOHENEGGER - A. LAMARRA - A. POLLOK - R. POZZO - H. PULTE - T. STURM, de Gruyter, Berlin/Boston 2023.

¹⁵ C.C.E. SCHMID, *Wörterbuch zum leichteren Gebrauch der Kantischen Schriften*, (hrsg.) N. HINSKE, Wissenschaftliche Buchgesellschaft, Darmstadt 2005; G.S.A. MELLIN, *Encyclopädisches Wörterbuch der kritischen Philosophie*, voll. I-III, Scientia, Aalen 1970-1971; R. EISLER, *Kant-Lexikon*, Olms, Hildesheim 1994; M. WILLASCHEK - J. STOLZENBERG - G. MOHR - S. BACIN (hrsg.), *Kant-Lexikon*, voll. I-III, de Gruyter, Berlin/Boston 2015.

¹⁶ H. VAHINGER, *Commentar zu Kants Kritik der reinen Vernunft*, Spemann, Stuttgart 1881.

¹⁷ G. MARTIN, *Allgemeiner Kant-Index*, de Gruyter, Berlin, 1967; A. ROSER - T. MOHRS - F.R. BÖRNCKE, *Kant-Konkordanz zu den Werken Immanuel Kants*, vol. I-III, Olms, Hildesheim 1992.

simile e ora utilizza il TEI¹⁸ e Norbert Hinske con il *Kant-Index* (costruito su TUSTEP)¹⁹, che forniscono entrambi l'accesso agli scritti di Kant in forma lemmatizzata con metadati e annotazioni semantiche che sono interoperabili e catturano il multilinguismo per quanto riguarda l'uso di Kant in greco, latino, tedesco e francese²⁰. Nel complesso, la ricerca sulle traduzioni di Kant in scritture non romane rimane un'area poco studiata²¹. Infine, un recente volume ha evidenziato il contributo della filosofia di Kant nel plasmare la ricerca contemporanea in filosofia della mente e nel chiarire alcuni dei problemi etici e pratici inerenti ai sistemi di intelligenza artificiale²².

L'allineamento delle *Druckschriften* e la loro traduzione completa in cinese²³ sono concepiti per aprire la strada a ulteriori progetti di allineamento, come nel caso dell'edizione Cambridge delle opere di Immanuel Kant²⁴, delle traduzioni russe coordinate dall'Istituto di Filosofia dell'Accademia delle Scienze Russa²⁵ e di altre traduzioni²⁶.

Proprio in questi mesi, un gruppo di ricercatori sta lavorando sul *backend* (livello di accesso ai dati) e sul *frontend* (livello di presentazione)

¹⁸ T. GREGORY - A. LAMARRA - E. PASINI - R. POZZO (curr.), *Lessico intellettuale europeo*, vol. I-CXXIX, Olschki, Firenze 1967-2022.

¹⁹ N. HINSKE (hrsg.), *Kant-Index*, Frommann-Holzboog, voll. I-XL, Stuttgart-Bad Cannstatt 1982-2017.

²⁰ H. HOHENEGGER, *From punchcards to digital archives: Forty years of ILIESI*, in R. POZZO - M. SGARBI (hrsg.), *Begriffs-, Ideen und Problemgeschichte im 21. Jahrhundert*, Harrassowitz, Wiesbaden 2011, 113-121.

²¹ R. POZZO, *Blick nach vorn. Kant-Übersetzungen und Korpora*, in *Archiv für Begriffsgeschichte Sonderheft 15* (2020), 323-334.

²² H. KIM - D. SCHÖNECKER (ed.), *Kant and Artificial Intelligence*, de Gruyter, Berlin/Boston 2022.

²³ LI QIULING, *Chuncui lixing pipan*, in *Kangde zhuzuo quanji*, voll. I-IX, China Renmin University Press, Beijing 2003-2019.

²⁴ P. GUYER - A. B. WOOD (ed.), *Cambridge Edition of the Works of Immanuel Kant*, voll. I-XVI, Cambridge University Press, Cambridge 1992-2016.

²⁵ B. TUSCHLING - N. MOTROSHILOVA (hrsg.), *Immanuel Kant Werke. Zweisprachige deutsch-russische Ausgabe*, voll. I-VI, RAS Institute of Philosophy, Moskau 1994-2020.

²⁶ G. SCHLÜTER - H. HOHENEGGER (hrsg.), *Kants Schriften in Übersetzungen*, in *Archiv für Begriffsgeschichte: Sonderheft 15*, Meiner, Hamburg 2020.

per creare un *corpus parallelo tedesco-cinese* dedicato alla filosofia di Kant e dunque anche agli scambi tra la terminologia filosofica occidentale e cinese²⁷. Di particolare interesse sono i *neologismi*, i prestiti dalle lingue occidentali alla triade cinese-giapponese-coreano²⁸. Nella preistoria delle *digital humanities*, i testi dovevano essere consultati manualmente prima di essere lemmatizzati e annotati. Oggi, lo spoglio viene eseguito da una macchina che lavora incorporando le parole in prestito. La parola di prestito incorporata nella lingua di partenza trova le sue corrispondenze incorporate nella lingua di arrivo.

La macchina rappresenta le parole in prestito nello spazio geometrico come vettori. È possibile eseguire algebricamente addizioni e sottrazioni di vettori per il rilevamento di novità su sequenze. La novità fondamentale va vista nell'interazione uomo-macchina per la progettazione partecipativa, con studi sugli utenti per la valutazione dell'impatto rispetto all'esperienza soggettiva degli utenti e alla quantità oggettiva del volume di dati.

La nuova edizione dell'Accademia sarà inoltre innovativa in quanto si propone di offrire accesso aperto a dei contenuti culturali, sperimentare nuovi modelli di business per l'industria editoriale, implementare un'analisi testuale multilingue dei testi kantiani, migliorare la conoscenza della grammatica occidentale in cinese, e spingere l'efficacia della traduzione automatica di testi letterari verso nuovi limiti.

5. Co-creazione

Come per le proteine, anche per le nuove edizioni di testi filosofici tempo e spazio non sono più rilevanti, perché ciascun ricercatore può

²⁷ R. POZZO - T. GATTA - H. HOHENEGGER - J. KUHN - A. PICHLER - M. TURCHI - J. VAN GENABITH, *Aligning Kant's Work and its Translations*, in D. FIŠER - A. WITT (ed.), *CLARIN. The Infrastructure for Language Resources*, de Gruyter, Berlin/Boston 2022, 690-708. DOI: 10.1515/9783110767377-029 [open access].

²⁸ T. GATTA, *Lo sviluppo del lessico filosofico nel cinese moderno*, Olschki, Firenze 2022.

accedere ovunque e a qualsiasi ora; gli utenti sono allo stesso tempo spettatori e ascoltatori, non più soggetti passivi ma attori, in grado di partecipare e contribuire. L'approccio dei *Large Language Models* richiede infatti la co-creazione, ovvero la partecipazione *open science* nell'uso dei dati, che sono conferiti o riutilizzati dagli archivi digitali mantenuti operativi dalle infrastrutture di ricerca. Nel caso delle scienze umane, in Europa sono CLARIN ERIC, DARIAH ERIC, EHRI, E-RIHS e OPERAS le infrastrutture di ricerca dedicate al dominio specifico dell'innovazione culturale. Sono infrastrutture che promuovono forme innovative di collaborazione tra scienziati e aiutano i ricercatori a produrre studi eccellenti, basati su dati digitali aperti, riutilizzabili, visibili e sostenibili, contribuendo così alla comprensione della vita culturale, economica, sociale e politica in Europa e oltre.

L'*Artificial Intelligence Act* dell'Unione Europea è la prima regolamentazione internazionale sull'intelligenza artificiale e contiene prescrizioni per i modelli di intelligenza artificiale ritenuti a rischio sistemico, basate su parametri quantitativi (capacità di calcolo) e qualitativi (supervisioni di comitati scientifici ed etici)²⁹. Nel mercato unico digitale europeo si prevede che le infrastrutture di ricerca delle scienze umane «rafforzino la ricerca sui contesti storici, sociali, economici, politici e culturali dell'Unione europea, fornendo dati e conoscenze a sostegno delle sue strategie»³⁰. Le scienze umane, del resto, producono non solo innovazione culturale³¹, ma anche innovazione sociale, la quale deve essere la «spina dorsale di tutte le politiche di ricerca europee», come ha detto chiaramente Marcelo Rebelo de Sousa, presidente della Repubblica portoghese, nelle osservazioni conclusive pronunciate in occasione della Conferenza di Lisbona, *Opening up to an Era of Social Innovation*

²⁹ <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence> (accesso 17.12.2023).

³⁰ <https://roadmap2021.esfri.eu/> (accesso 17.12.2023).

³¹ R. POZZO - A. FILIPPETTI - M. PAOLUCCI - V. VIRGILI, *What Does Cultural Innovation Stand for? Dimensions, Processes, Outcomes of a New Innovation Category*, in *Science and Public Policy* 47 (2020) 425-433. DOI 10.1093/scipol/scaa023 [open access].

Conference, del 28 ottobre 2017³². E non dimentichiamo che l'innovazione sociale si verifica quando un servizio risponde positivamente ai problemi della società – vale a dire, l'innovazione risolve il problema? Costa il giusto? È universalmente accettata?³³ Questo intervento può essere allora una buona occasione per riflettere sulle pratiche editoriali di questi ultimi anni e sulle vie da percorrere per rendere la consultazione non solo più immersiva, ma anche più partecipata, dunque in contesti di co-creazione.

³² EUROPEAN COMMISSION, *Opening up to an Era of Social Innovation*, Publication Office of the European Union, Luxembourg 2019, DOI: 10.2777/11297 [open access].

³³ F. MOULAERT - A. MEHMOOD - D. MACCALLUM - B. LEUBOLT, *Social Innovation as a Trigger for Transformations*, Publication Office of the European Union, Luxembourg 2017, DOI: 10.2777/68949 [open access].

INTELLIGENZA ARTIFICIALE COME “TECNOLOGIA DEL SÉ”

Andrea Galluzzi

Abstract

Le tecnologie digitali influiscono sull'identità personale: l'IA può essere un mezzo al servizio della persona, ma occorre riflettere sull'armonizzazione tra la costruzione individuale di un sé e la fragilità di un noi impreparato ad affrontare questa evoluzione. Prendendo spunto dal concetto di *tecnologie del sé* di M. Foucault, mettiamo qui in luce alcune connessioni tra filosofia, sociologia e antropologia nel contesto attuale, in cui l'IA sta facendo evolvere il concetto di macchina da “strumento” a “compagno di lavoro”.

Parole chiave: IA, Identità personale, Tecnologia del sé, M. Foucault, Retroazione.

Digital technologies affect personal identity: AI can be positioned as a means at the service of the person, but it is necessary to reflect on the harmonization between the individual construction of a self and the fragility of an us unprepared to face this evolution. Inspired by the concept of technologies of the self by M. Foucault, this study intends to highlight some connections between philosophy, sociology, and anthropology in the current context, in which AI is evolving the concept of the machine from “tool” to “workmate”.

Keywords: AI, Personal identity, Technologies of the self, M. Foucault, Feedback.

1. Tra realtà e simulazione. I nuovi mediatori digitali

Scienza, tecnica e tecnologia plasmano il nostro mondo e – soprattutto dall'avvento dell'era digitale – modificano in maniera sempre

più pervasiva i nostri ritmi di vita, le nostre abitudini, il nostro modo di pensare e agire e le nostre relazioni sociali. L'avvento dell'Intelligenza Artificiale (IA) ha cambiato ulteriormente questo scenario, superando progressivamente i confini della ricerca e della sperimentazione per entrare nel mondo reale e divenire una tecnologia di consumo. I modi in cui i dispositivi dotati di IA agiscono nella sfera umana sono innumerevoli e nel tempo il loro impatto si è progressivamente avvicinato alla sfera individuale, fino alla possibilità di assumere in maniera crescente il ruolo di mediatori fra l'individuo e il mondo circostante; quasi una "guida" in grado di farci orientare nel complesso universo digitale. Ma in che modo e in che senso agiscono queste tecnologie? E qual è il loro ruolo nella nostra relazione con gli altri?

Fra le grandi possibilità aperte dalle tecnologie informatiche ci sono la modellizzazione e simulazione di fenomeni reali. La simulazione è un elemento fondamentale nella progettazione e nell'ingegneria: è un'operazione che fornisce i dati necessari a prevedere l'insorgenza di problemi strutturali e permette di pensare in anticipo le strategie da attuare per cercare di risolverli. L'evoluzione dei sistemi di calcolo ha permesso via via la modellizzazione di sistemi sempre più complessi e sta trovando applicazione ovunque si riesca a raccogliere dati utili. La simulazione computazionale dei fenomeni ha permesso, ad es., di sostituire in tutto o in parte l'esperimento fisico, attuando una sorta di nuova rivoluzione scientifica in grado di accelerare ulteriormente, anche grazie all'IA, i processi di ricerca e sviluppo¹. Le sempre maggiori capacità di calcolo disponibili e le crescenti capacità di gestione dei servizi di rete hanno ampliato questo scenario, aprendo la strada a una più ampia fruizione

¹ A titolo di esempio: nel 2022 DeepMind, azienda controllata da Google, ha comunicato di essere riuscita a predire computazionalmente la struttura tridimensionale di oltre 200 milioni di proteine, superando in pochi mesi decenni di ricerca scientifica infruttuosa e aprendo la strada alla cura innovativa di molte patologie. Le tecniche di modellizzazione digitale portano moltissimi vantaggi nel mondo della progettazione, della produzione, dell'analisi preventiva dei fenomeni. Cf. D. HASSABIS, *AlphaFold reveals the structure of the protein universe*, 28 luglio 2022, <https://deepmind.google/discover/blog/alphafold-reveals-the-structure-of-the-protein-universe> (accesso 5.12.2023).

di servizi digitali innovativi da parte di un numero crescente di individui.

Il carattere di queste simulazioni, tutte inizialmente raggruppate sotto il nome di *realtà virtuale*, nel corso del tempo ha assunto maggiore spessore, prendendo in qualche modo forza e solidificandosi attorno a concetti più definiti, come nel caso dei *metaversi*, che possiamo qui intendere come simulazioni di ambienti reali o immaginari; o il caso dei *gemelli digitali*, cioè trasposizioni di oggetti reali in spazi virtuali, creati digitalmente, al fine di simulare comportamenti e interazioni o testare scenari senza necessariamente coinvolgere l'oggetto fisico che essi rappresentano².

La creazione di *gemelli digitali* non è circoscritta agli oggetti inanimati, ma può riguardare anche il mondo organico, permettendo di modellare e analizzare il comportamento e le caratteristiche di organismi viventi in contesti simulati. Grazie al proliferare di dispositivi indossabili³, in grado di monitorare in maniera non invasiva molti nostri parametri fisiologici, si è aperta la possibilità di vedere rappresentato in dettaglio anche il nostro sé biologico, aprendo scenari inediti per la medicina digitale. Questo settore di ricerca si propone infatti di creare duplicati virtuali del corpo umano, da studiare per prevenire, diagnosticare o trattare le malattie in maniera più efficiente, o su cui testare gli interventi terapeutici in tempi brevi, a vantaggio della nostra salute.

² Sono attualmente disponibili online piattaforme (come "gemelli digitali" di Azure) che abilitano la creazione di modelli digitali gemelli per interi contesti, quali edifici, fabbriche, reti energetiche, infrastrutture ecc. Questi modelli digitali sono fondamentali per ottenere informazioni che possano guidare l'ottimizzazione dei prodotti, migliorare la gestione delle operazioni, ridurre i costi, ma anche trasformare l'esperienza di utilizzo dei prodotti stessi. <https://azure.microsoft.com/it-it/products/digital-twins> (accesso 5.12.2023).

³ I dispositivi dedicati alla rilevazione dei parametri fisiologici più comuni sono i già diffusissimi *smartwatch*, *fitness-band* e *smart-rings*, ma anche auricolari *wireless* con funzioni sanitarie e tessuti elettronici in grado di essere utilizzati in varie forme di abbigliamento "smart". Sulla classificazione di questi strumenti come veri dispositivi medici o come "gadget sanitari" si dibatte da tempo, ma la diffusione di queste soluzioni è in continua crescita e il mercato in forte espansione, segno di una sensibilità crescente sugli aspetti legati all'assistenza sanitaria nella sfera personale.

L'IA sta entrando anche nel settore sanitario attraverso varie vie di accesso. Gli attuali sistemi in grado di processare il linguaggio naturale possono infatti trovare efficacemente impiego nella creazione e comprensione della documentazione clinica, con la capacità di elaborare dati e note cliniche non strutturati; oppure conversando con i pazienti per offrire vari tipi di supporto. Il processo di sviluppo delle IA in campo sanitario è destinato inevitabilmente a crescere, viste la complessità degli aspetti clinici e la mole sempre più massiccia di dati a disposizione. Solo la rapida capacità predittiva di una IA potrà infatti analizzare in tempo reale i dati dei pazienti e le correlazioni in grado di identificare l'imminenza di eventi clinicamente rilevanti, allertando in tempo una specifica assistenza sanitaria.

Queste possibilità stanno aprendo la strada all'adozione di IA personali capaci di prendersi cura di noi non solo sotto l'aspetto medico, ma anche per le nostre attività e i nostri interessi. I dati che ci rappresentano e che sono memorizzati sui server online, infatti, possono essere interpretati come una copia informativa di noi stessi che abita l'interno del mondo digitale e che ci offre nuove modalità di rapportarci ad esso attraverso nuove forme di mediazione. Il nostro *gemello digitale* potrà fare le scelte per noi e saprà muoversi meglio di noi in quel mondo, perché dentro la sua natura. Si potrebbe dire che il gemello digitale è il vero soggetto che agisce nel mondo digitale. Analizzando tutti i nostri dati (dalle analisi cliniche alle abitudini alimentari, dall'umore quotidiano agli interessi personali) il nostro *gemello digitale* sarà in grado di supportarci come un mentore e tutelare la nostra salute più efficacemente di un medico personale o un tutore onnipresente.

Questa sorta di personificazione dei nostri alter-ego digitali ci porta a fare qualche ulteriore riflessione sul rapporto con le tecnologie attuali, e in particolare con le IA. Possiamo paragonare questi dispositivi a una sorta di "scudo" o, meglio, a un *guscio tecnologico* proattivo attorno alla persona. 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, interponendosi così fra il soggetto e il mondo circostante, aprendo la possibilità di costruire nuove definizioni sociali di sé.

2. La costruzione delle identità personali nell'*Onlife*

I domini di conoscenza del mondo digitale sono andati progressivamente ampliandosi nel corso del tempo e hanno portato a formalizzare concetti via via più definiti al suo interno. Il *virtuale* ha acquisito una identità dai contorni più netti, assumendo un nuovo carattere ed evolvendo verso la definizione di *metaverso*, cioè una realtà parallela, autonoma, alla quale poter accedere attraverso opportune interfacce. L'idea del *metaverso* nasce come allontanamento dalla realtà materiale, ma sta crescendo nella direzione di essere sempre più *parte* integrante del *tutto* nel quale viviamo: un mondo nel quale si va assottigliando, fino a eluderlo, il confine fra il reale e il virtuale.

Per cercare di capire come queste tecnologie stiano influenzando la sfera umana, dobbiamo provare a ripercorrere la storia recente. Nella seconda metà degli anni '90 del secolo scorso, con l'avvento di internet, hanno iniziato a farsi strada i primi tentativi di connessione in rete a livello sociale, che hanno trovato espressione sia sotto forma di simulazioni virtuali (come *Second Life*, fondato nel 2003) sia come social network (diffusisi massivamente con *Facebook* a partire dal 2004). Il verificarsi di questi eventi ha rappresentato l'inizio di un nuovo momento storico: le tecnologie dell'informazione hanno iniziato a farsi strada nel campo della formazione delle nostre identità sociali. I nuovi media non hanno solo introdotto nuove abitudini o stili di vita – cosa che avviene normalmente quando vengono introdotti nuovi apparati tecnologici di uso comune – ma hanno dato a ognuno l'opportunità di accedere agli strumenti utili per potere costruire una rappresentazione di sé al mondo.

Nel processo di definizione della propria identità il contesto sociale gioca un ruolo rilevante: il modo in cui ci vedono gli altri influisce infatti su ciò che percepiamo di noi stessi. Questa prospettiva sociale è ampiamente modellata dalle tecnologie dell'informazione, soprattutto dai social media, che hanno un impatto significativo in questa dinamica di auto-rappresentazione e costruzione di sé in rete. La questione identitaria online non riguarda una ristretta cerchia di persone, ma la quasi

totalità di chi è connesso a internet⁴ e ovviamente innesca comportamenti che vanno oltre le appartenenze culturali e di genere. Il fenomeno della trasformazione identitaria attraverso i social network non riguarda solo i giovanissimi, ma attraversa trasversalmente più generazioni e richiede una maggiore consapevolezza riguardo le potenzialità e i rischi presenti in queste nuove modalità di comunicare.

Un altro mutamento significativo nell'evoluzione della nostra interazione con la tecnologia a livello sociale è rappresentato dal fenomeno (non ancora pienamente affermatosi massivamente) dei metaversi, in cui si immagina che i nostri *avatar* potranno dare origine a nuove forme di narrazione della nostra vita virtuale, sempre più mescolata con quella reale.

Tutte queste possibilità incontrano in maniera sempre più pervasiva le potenzialità dell'IA, aprendo ad ulteriori espansioni del sé. L'allargamento dei confini personali è frutto della crescente consapevolezza che i nostri dati, la loro strutturazione e gli algoritmi che li processano ci portano in qualche modo a prolungare la nostra corporeità, a dilatarla senza un confine preciso e a veicolarla verso gli altri attraverso la rete e le sue interfacce. È intuitivo pensare come le regole di questa estensione potranno essere co-gestite dai nostri avatar digitali, veicolatori delle nostre identità nel mondo online⁵.

⁴ Stime recenti dicono che 4,76 miliardi di persone sono attive sui social a fronte di 5,16 miliardi connesse in rete, e ad una popolazione globale di 8,01 miliardi di individui. Cf. S. KEMP, *Digital 2023: a global overview report*, 26 gennaio 2023, <https://datareportal.com/reports/digital-2023-global-overview-report> (accesso 5.12.2023).

⁵ Su questo fronte le nuove IA generative, impostesi massivamente sul mercato a partire dal novembre 2022 con la diffusione di ChatGPT, acquisteranno sempre maggiore rilevanza. L'annuncio di OpenAI sull'arrivo delle versioni personalizzate di ChatGPT, ad appena un anno dal suo debutto, va infatti nella direzione di un ulteriore incremento tecnologico: ognuno potrà creare la propria IA e configurarla a piacimento. Se queste innovazioni guardano già alle imprese e ai servizi loro annessi, in un prossimo futuro toccheranno anche la sfera personale, offrendo nuove opportunità di personalizzazione grazie alle sempre più avanzate capacità delle IA generative.

Le possibilità aperte dalle nuove tecnologie suscitano il bisogno di nuovi riferimenti concettuali e nuove categorie di pensiero. Dall'inizio dell'era dell'informatizzazione di massa (che possiamo stabilire con vari criteri, ma in questo contesto preferiamo inquadrarla con l'avvento dei personal computers e, più tardi, di internet) la cultura digitale si è infatti evoluta facendo emergere concetti nuovi. L'autore e ricercatore statunitense Ed Finn (1980-), ad es., utilizza il termine *persona algoritmica*⁶ per indicare un'entità definita da una distribuzione di dati e strutture sparsi fra un dispositivo e l'altro e correlata alla persona reale, alla quale i dati si riferiscono e sulla quale retroagiscono fornendo informazioni personalizzate. Finn parla di una "ingegnerizzazione" della nostra intimità per indicare quanto le tecnologie siano già entrate nella sfera dell'interiorità:

«la pervasiva e profonda raccolta di dati svolta da aziende come Google ci viene offerta non come una nuova forma di coscienza ma come una serie di strumenti per la conoscenza di sé [che vanno dal conoscere precisamente i propri ritmi circadiani alla tenuta del proprio passo durante la corsa, ai programmi per la settimana]. È una forma di propriocezione algoritmica, un'espansione del sé, resa possibile dai sensori fisici e cognitivi [...]»⁷.

Per indicare la saldatura fra la vita all'interno e al di fuori della rete globale, il filosofo Luciano Floridi (1964-) ha coniato il noto termine *Onlife*⁸, un concetto sempre più attuale perché cresce con la consapevolezza della nostra natura informazionale. Floridi evidenzia che gli individui possono essere considerati come *organismi informazionali interconnessi (infor)*⁹, che condividono il proprio ambiente con altri agenti informazionali e altri artefatti ingegneristici. Sebbene questa definizione

⁶ Cf. E. FINN, *What algorithms want. Imagination in the age of computing*, Mit Press, Cambridge 2017, 76.

⁷ *Ibid.*, 77 (traduzione mia).

⁸ Cf. L. FLORIDI, *The 4th revolution. How the infosphere is reshaping human reality*, Oxford University Press, New York 2014.

⁹ Cf. *Ibid.*, 94-96.

non esaurisca certamente l'insondabile complessità dell'essere umano, essa suscita comunque l'idea del peso assunto dalla teoria dell'informazione e della complessità dell'*infosfera*¹⁰ che abitiamo. Insistendo sulla nostra natura informazionale, Floridi fornisce ulteriori precisazioni: il fatto che i nostri corpi, nelle loro componenti ultime, siano costituiti da informazioni, non giustifica infatti un riduzionismo dualista fra mente e corpo. Per evitare questa deriva, il filosofo invita a ragionare più in termini di un monismo basato sulla nozione di *stato* e sulla mutazione della sua forma (come l'acqua che, pur rimanendo se stessa, si può presentare in forma solida, liquida o gassosa). In questa prospettiva la dicotomia tra mente e corpo si riduce, in quanto entrambe le componenti possono essere considerate come due differenti configurazioni informazionali, due facce della stessa medaglia, o meglio: «due diversi stati di una componente informativa sottostante»¹¹.

Data questa particolare natura informazionale, la costruzione dell'identità personale avviene nello scambio fra informazioni interiori ed esteriori, cioè nella relazione fra ciò che percepiamo internamente a noi stessi e l'*infosfera* che avvolge il mondo. Argomentando su come quest'ultima stia cambiando la concezione che abbiamo di noi stessi, Floridi mette in luce l'importanza della relazione fra la componente individuale (cioè la nostra identità personale, che identifica chi siamo realmente) e quella sociale, che influenza inevitabilmente la prima e si riflette su ciò che noi stessi pensiamo di essere.

La fioritura dell'individuo è legata all'equilibrio fra queste due componenti e al modo in cui esse si sostengono reciprocamente, senza dimenticare che la nostra personalità sociale è costruita anche dal

¹⁰ Il termine *infosfera* è un neologismo coniato negli anni '70 del Novecento derivato dal termine *biosfera*. L'*infosfera* «denota l'intero ambiente informazionale costituito da tutti gli enti informazionali, le loro proprietà, interazioni, processi e reciproche relazioni [...]. [Ma] è un concetto che può essere utilizzato anche come sinonimo di realtà, una volta che questa sia interpretata in termini informativi. In questo caso, il suggerimento è che ciò che è reale è informazionale e ciò che è informazionale è reale». *Ibid.*, 40-41 (traduzione mia).

¹¹ *Ibid.*, 71 (traduzione mia).

pensiero altrui, oggi veicolato pesantemente dai social media¹².

In sintesi: le tecnologie digitali, e in particolare le IA, sono entrate nella dinamica costruttiva delle nostre identità personali – che si sviluppano attraverso le interazioni sociali nel contesto socio-culturale in cui viviamo – agendo come intermediari tra noi e la grande *infosfera* che è il mondo in cui abitiamo. È in questo contesto che Floridi richiama un concetto significativo ai fini della nostra riflessione:

«Il sé è visto come un sistema informativo complesso, composto da attività di coscienza, ricordi o narrazioni [...]. E poiché le tecnologie dell'informazione (*Information and Communication Technologies*) possono influenzare profondamente questi schemi informativi, esse sono effettivamente potenti *tecnologie del sé*»¹³.

Il termine *tecnologie del sé* è stato coniato dal filosofo francese Michel Foucault (1926-1984) nel contesto della sua ricerca sul modo in cui gli esseri umani sviluppano la conoscenza di sé stessi. Alcune considerazioni del filosofo francese offrono ulteriori spunti per la nostra riflessione e permettono di estenderla in un orizzonte più ampio. Le implicazioni pratiche di queste applicazioni tendono infatti a retroagire sempre più profondamente sul nostro modo di intendere l'esistenza e agiscono nell'intreccio fra individuo e collettività.

3. Il sé e il noi attraverso le tecnologie

L'idea intorno alle *tecnologie del sé*¹⁴ è stata maturata da Michel Foucault nell'ultima fase della sua vita e va contestualizzata all'interno

¹² Cf. *Ibid.*, 60.

¹³ *Ibid.*, 69 (traduzione e corsivo miei).

¹⁴ Nel testo di Foucault i termini *tecnologie del sé* e *tecniche del sé* sono utilizzati come sinonimi, in modo interscambiabile. Nel contesto della nostra trattazione hanno invece significati diversi, anche se correlati fra loro: con il termine *tecniche del sé* vogliamo qui indicare le "procedure" o le "modalità di attuazione" o le "pratiche" orientate al sé, mentre con il termine *tecnologie del sé* facciamo riferimento ai dispositivi, frutto dell'ingegno umano e della scientificizzazione delle tecniche, che vengono utilizzati nel modo descritto

di una riflessione più ampia del filosofo francese, legata all'investigazione delle pratiche attraverso le quali gli individui ricercano il benessere, la perfezione, la saggezza o, più in generale, la conoscenza (o svelamento) di sé. Con questo termine il filosofo francese intendeva indicare tutte quelle pratiche e mezzi in grado di permettere agli individui di eseguire una trasformazione di se stessi allo scopo di raggiungere uno stato superiore di felicità e realizzazione.

Benché risalga ai primi anni Ottanta del secolo scorso, la riflessione filosofica di Foucault sulle tecniche per la formazione del sé si adatta bene anche al nostro tempo perché, indipendentemente dal grado di complessità tecnologica raggiunto dalla nostra società, essa ragiona sulla genesi culturale che ha plasmato il pensiero moderno e che sostiene lo sviluppo tecnologico in corso. La riflessione si articola a partire dai due principi morali più importanti del pensiero greco: *conosci te stesso* e *prenditi cura di te*: «nei testi greci e romani – sottolinea Foucault – il precetto di conoscere se stessi è sempre associato a quello della cura di sé, ed era proprio questo bisogno di prendersi cura di sé che rendeva operativa la massima delfica»¹⁵. Foucault fa notare che nella cultura occidentale si è invece assistito ad una graduale inversione gerarchica fra i due principi, ponendo la conoscenza di sé come principio fondamentale e lasciando nell'ombra il secondo. L'obiettivo del filosofo francese è quello di fare riscoprire il rapporto originario fra i due principi, analizzare i motivi storici di questa inversione e mettere in evidenza le pratiche (le *tecniche del sé*) in grado di favorire una rivalutazione dell'individuo e del proprio equilibrio esistenziale.

Nel quadro della nostra analisi possiamo riformulare il pensiero di Foucault inserendo fra le tecniche per la conoscenza e formazione del sé anche l'utilizzo delle tecnologie digitali orientate alla persona. Nell'ottica della *cura di sé* si possono infatti annoverare i dispositivi indossabili dedicati alla analisi dei parametri fisiologici cui abbiamo

in queste righe.

¹⁵ M. FOUCAULT, *Tecnologie del sé*, in: L. H. MARTIN - H. GUTMAN - P. H. HUTTON (curr.), *Tecnologie del sé. Un seminario con Michel Foucault*, Bollati Boringhieri, Torino 2005, 11-47, 15.

accennato prima; mentre per la *conoscenza di sé* possiamo considerare le influenze dell'*infosfera* sociale nel modo esposto precedentemente. Il tutto considerando anche l'azione coadiuvante delle IA.

Per cercare di comprendere cosa è in gioco nel considerare l'IA come *tecnologia del sé*, e come questo vada ad impattare sulla dimensione sociale, proverò a partire dal concetto di *azione* e a tracciare una linea di pensiero che possa portare a definire una risposta plausibile. La parola *agire* (dal latino *agere*) indica l'atto di mettere in movimento qualcosa. Nel suo senso più generale significa iniziare, prendere iniziativa, dare cominciamento a qualcosa (come indica la parola greca *archein*)¹⁶. Il tema dell'azione è importante perché serve a fare chiarezza sulle cause dei comportamenti individuali e dei fenomeni collettivi, e a cercare spiegazioni sulle logiche che ne sono alla base. Ad agire però non sono solo gli esseri umani o solo ciò che è naturale, ma anche le macchine, i sistemi automatici, gli algoritmi e le IA, in un intreccio di attività, correlazioni e interazioni in cui gli uni incidono sugli altri, innescando continue trasformazioni causate dall'agire altrui. Tutto si riflette comunque sulla sfera umana.

Le nostre azioni sono legate a un insieme di condizionamenti culturali, alle nostre scelte personali e anche – come insegnano le scienze sociali – alla razionalità che le sottende. Fra i vari tipi di razionalità che possono guidare l'agire, la razionalità tecnica sembra giocare oggi un ruolo molto importante nello sviluppo della nostra personalità e del nostro modo di rapportarci agli altri, proprio a causa degli apparati tecnici che noi stessi abbiamo messo in atto e che oggi pervadono anche gli ambiti più vitali. La tecnica infatti, oltre ad indicare il modo in cui compiamo azioni per raggiungere scopi, indica anche quella forma della razionalità umana che presiede all'impiego degli strumenti, che è orientata all'efficienza, alla concretezza pratica, all'ottimizzazione, alla funzionalità. La tecnica insomma ci appartiene, fa parte di noi, è una componente antropologica peculiare che permette di identificare la nostra condizione come una

¹⁶ Cf. H. ARENDT, *Vita activa. La condizione umana*, Bompiani, Milano 2011, 128-129.

*condizione tecno-umana*¹⁷, che si esprime e si riflette nelle nostre azioni e negli apparati che mettiamo in esercizio.

Su queste premesse la relazione uomo-macchina arriva a strutturarsi secondo una dinamica per descrivere la quale possiamo ricorrere a un termine mutuato dalla teoria dei sistemi e della cibernetica: la retro-azione (o *feedback*). In un sistema dotato di ingressi e uscite la retroazione fa sì che le uscite possano tornare ad influenzare il sistema stesso, modificandone lo stato. Questa influenza si può attuare sia nel senso di una forma di regolazione (come, ad esempio, si controlla il livello dell'acqua nei bacini delle centrali idroelettriche analizzando la differenza fra il flusso in entrata e quello in uscita) sia nel senso di un incremento sempre più veloce e dilatato, come avviene nelle reazioni a catena o nei fenomeni fisici di risonanza.

Le macchine e gli automatismi non sono dotati di razionalità, ma agiscono, e il loro agire influisce sulle nostre stesse azioni. In altre parole: le tecnologie retro-agiscono sulla sfera umana secondo modalità che tendono a sfuggire al nostro stesso controllo. La crisi ambientale è un esempio degli effetti della retroazione incontrollata dello sviluppo industriale sull'ecosistema globale, ma questi concetti possono essere trasposti anche in ambito sociale, considerando gli effetti che alcune innovazioni hanno avuto e stanno continuamente avendo (nel bene e nel male) sul nostro modo di vivere e sulle nostre abitudini, sia sul piano individuale che collettivo. Malgrado la accresciuta sensibilità su questi temi e l'attuazione di adeguate contromisure, ci troviamo continuamente davanti a nuovi problemi. Le cause di questo fenomeno sono innumerevoli, ma si possono riassumere in un unico principio: la rapidità con cui si innovano le tecnologie è superiore alla nostra capacità di adattamento¹⁸.

¹⁷ Per approfondire: P. BENANTI, *La condizione tecno-umana. Domande di senso nell'era della tecnologia*, EDB, Bologna 2016; C. CALTAGIRONE - L. CUCURACHI (curr.), *La condizione tecno-umana tra eccesso ed eccedenza*, Morcelliana, Brescia 2021.

¹⁸ Avere capacità di adattamento non significa adeguarsi passivamente alla nuova situazione di esistenza, ma sapere modificare il proprio comportamento o le proprie abitudini

Il processo di retroazione innescato dalle tecnologie è del tutto naturale e ci caratterizza propriamente come specie umana sin da quando abbiamo iniziato a controllare il fuoco e lavorare la selce. La rivoluzione industriale e quella digitale non sfuggono a questo principio, ma hanno impresso a questo processo una accelerazione progressiva che ha portato con sé anche una certa forma di destabilizzazione esistenziale. Ogni nuova invenzione, nel bene e nel male, ha condizionato la nostra storia, ma mai come negli ultimi decenni abbiamo assistito all'introduzione di invenzioni in grado di suscitare fenomeni di massa e influire trasversalmente sulle nostre abitudini, sui riferimenti culturali e di costume, ma anche, più profondamente, sul nostro modo di vedere il mondo, e sul modo di costruire la nostra identità. E questa – come abbiamo ricordato – è costruita socialmente, cioè grazie alle azioni, interazioni e relazioni con gli altri in un contesto oggi fortemente caratterizzato dallo sviluppo tecnologico.

L'assunzione di tecnologie basate su IA non sfugge al principio di retroazione, perciò esse possono certamente porsi come potente mezzo al servizio della persona, ma occorre riflettere sulla delicata armonizzazione tra la costruzione individuale di un sé e la fragilità di un *noi* ancora impreparato ad affrontare le sfide e le conseguenze di questa evoluzione. In sintesi, tornando ai principi esposti all'inizio di questo paragrafo, l'ambivalente potenziamento indotto dalle tecnologie digitali può agire nella direzione di rafforzare cura e conoscenza di sé, ma anche ribaltarsi, enfatizzando la chiusura della persona su se stessa. Osservando infatti l'evolversi della diffusione dei sistemi informazionali digitali e le ripercussioni su individuo e società, ci si può accorgere di come questi abbiano in un certo modo favorito l'emancipazione del soggetto, ma anche un generale disorientamento valoriale.

Gli effetti delle tecnologie sulla sfera umana, oltre ad evidenti benefici, sono infatti spesso orientati in direzione di un potenziamento del sé e, anche pur volendo valorizzare la persona, rischiano di risolversi in una

per attuare nuove regolazioni, in armonia con le nuove circostanze.

involuzione del soggetto all'interno del proprio *guscio tecnologico*.

A rafforzare questa idea contribuisce il fenomeno delle nuove applicazioni che consentono di utilizzare le IA allo scopo di compagnia e supporto, la cui popolarità è in aumento. Una tendenza tecnologica già in atto è infatti quella che vede utilizzare le IA per simulare persone con le quali colloquiare, giocare, condividere contenuti o ottenere sostegno morale o psicologico. Le caratteristiche dell'IA personale vengono decise dagli utenti, che creano così compagni "su misura", senza i problemi e le ansie che si possono generare nei rapporti reali, visto che questi sono inevitabilmente condizionati dalla libertà e personalità altrui¹⁹. Scenari come questo ci portano a comprendere la necessità di riuscire a individuare forme di analisi e azione sociale che permettano di valorizzare il concetto di relazione e capire come questo possa essere messo a frutto nel contesto della nostra trattazione.

4. Dal sé al noi. Le relazioni umane oltre il *guscio tecnologico*

Riflettere sulla nostra esistenza mediata dalle IA ci spinge a esaminare non solo l'aspetto tecnologico in sé, ma anche – sul piano antropologico – le dinamiche comportamentali e sociali che derivano dalla mediazione del *guscio tecnologico* che sembra avvolgere ogni individuo. In questa prospettiva, il principio secondo il quale a una evoluzione tecnologica deve corrispondere una altrettanto solida crescita umana diventa sempre più attuale e urgente. Tutto ciò si traduce nel cercare di trovare una prospettiva per dare significato alle nostre azioni nel contesto mediato dalle tecnologie, in modo da valorizzare la peculiarità

¹⁹ Fra le varie applicazioni dedicate a questo scopo, quella maggiormente diffusa è Replika (Replika.com), trainata al successo dalla ricerca di compagnia e supporto durante la pandemia da COVID-19, oltre che dall'interesse degli utenti per il supporto psicologico e l'auto-miglioramento. Da qualche tempo è possibile utilizzare anche la realtà aumentata per visualizzare i compagni digitali nel proprio ambiente circostante, così da simulare un vero e proprio incontro.

relazionale dell'essere umano come persona capace di farsi responsabile del proprio agire. A tale scopo possiamo riprendere i concetti a cui abbiamo accennato sopra, ampliandoli con il contributo di alcuni filosofi che hanno approfondito il nesso fra *azione* e *relazione*. Hannah Arendt (1906-1975), ad es., analizzando la condizione umana, identifica nell'azione e nel discorso (cioè la capacità di articolare il linguaggio) ciò che caratterizza propriamente l'essere umano, mettendo in evidenza il ruolo degli "altri":

«L'azione, diversamente dalla fabbricazione, non è mai possibile nell'isolamento; essere isolati significa essere privati della facoltà di agire. Azione e discorso necessitano della presenza degli altri, allo stesso modo in cui la fabbricazione necessita della natura e dei suoi materiali, e di un mondo in cui collocare il prodotto finito. La fabbricazione è circondata dal mondo con cui è in costante contatto; l'azione e il discorso sono circondati dall'intreccio e dalle parole di altre persone con cui sono in costante contatto»²⁰.

Rileggendo l'intreccio delle relazioni umane nelle narrazioni dell'antichità, Arendt vede emergere il "coraggio" come elemento originario senza cui l'azione e il discorso non sarebbero possibili²¹. Occorre cioè esercitare coraggiosamente la capacità di uscire da sé e rivelarsi per quello che si è, con pregi e difetti. Azione e discorso hanno infatti una capacità rivelativa e questa emerge quando si è con gli altri: «agendo e parlando gli uomini mostrano chi sono, rivelano attivamente la loro identità personale, e fanno così la loro apparizione nel mondo umano [...]»²². Seguendo il pensiero di Hannah Arendt, il desiderio di sfuggire alla condizione umana seguendo l'idea del potenziamento e del superamento dei propri limiti attraverso le tecnologie, mettendo in secondo piano le relazioni con gli altri, ci porterebbe sulla via che conduce alla condizione di schiavitù. Questa subordinazione non si esprimerebbe tanto nei confronti delle macchine, quanto delle nostre stesse competenze,

²⁰ ARENDT, *Vita activa*, cit., 137.

²¹ Cf. *Ibid.*, 136.

²² *Ibid.*, 130.

e finirebbe nel farci perdere il controllo sul potere esercitato da ogni dispositivo tecnicamente possibile, del quale rischiamo di essere alla mercé. Solo l'uomo "al singolare" potrebbe trarre giovamento da una chiusura nel suo – diremmo qui – *guscio tecnologico*, ma nella pluralità che siamo e che ci caratterizza come esseri umani è necessario dare un senso al nostro agire in modo plurale e reciproco²³.

Il filosofo Adriano Fabris (1958-) pone un legame ancora più stretto fra "azione" e "relazione", identificando l'azione come il "motore" della relazione e attribuendo all'agire un significato sociale: «Agire è il modo in cui la relazione si fa come tale. Agire è il centro dinamico di ogni rapporto. Perciò [...] possiamo parlare in proposito di (rel)azione o, meglio, di relAzione»²⁴. L'agire, per Fabris, è «il modo in cui qualcosa si esprime in forma relazionale»²⁵ e la relazione, a sua volta, è costitutivamente anche la condizione di possibilità per agire eticamente. In questa prospettiva, la dinamica relazionale non retroagisce sul singolo in modo limitante ma, al contrario, «l'etica della relazione è chiamata a sviluppare le possibilità dell'agire, invece che a ridurle»²⁶.

Considerando la nostra natura relazionale, libertà, responsabilità e potenza (intesa come spinta dell'eccentricità umana) si trovano inscindibilmente legate. Si tratta però di evitare di consegnare questi elementi al monopolio del sistema tecnico e dei suoi apparati, con il rischio di trasformare la potenza in desiderio di onnipotenza. È questa la prospettiva messa ampiamente a tema dal sociologo Mauro Magatti (1960-) il quale propone una «rinnovata riflessione sulla potenza che, come singoli e come collettività, siamo diventati capaci di produrre»²⁷: per potersi salvare dal desiderio di onnipotenza – cioè il pericolo di essere assoggettati dall'illusione di dominare la realtà – occorre sapere accettare la

²³ Cf. *Ibid.*, 3-4.

²⁴ A. FABRIS, *RelAzione. Una filosofia performativa*, Morcelliana, Brescia 2016, 119.

²⁵ *Ibid.*, 126.

²⁶ *Ibid.*, 117.

²⁷ M. MAGATTI, *Oltre l'infinito. Storia della potenza dal sacro alla tecnica*, Feltrinelli, Milano 2018, 16.

condizione del limite, e dunque sapere rinunciare a qualcosa.

Il fatto che non ci si possa salvare senza qualche rinuncia è un principio che in qualche modo hanno messo in evidenza vari autori, tutti mossi dal medesimo obiettivo: tentare di capire come contrastare il potere fuori controllo delle applicazioni della tecnica. Fra questi autori si possono annoverare Romano Guardini (1885-1968) e Giorgio Agamben (1942-), entrambi citati da Magatti per spiegare il concetto di *deponenza*²⁸. Con questo termine Magatti intende descrivere la condizione umana come un intreccio di attività e passività in cui si contrappongono potenza e sovranità. Esse non possono coesistere, ma solo essere presenti in maniera mutuamente escludente: «Essendo impastata con la vita, se vuole rimanere umana, la potenza (religiosa, politica, tecnica) deve rinunciare al mito della sovranità»²⁹. Si tratta, in sintesi, di mettere da parte l'idea di onnipotenza indotta dalla tecnica per riprendere il controllo della potenza, pena il fallimento umano totale, individuale e collettivo. Occorre in pratica sapere anche rinunciare a qualcosa di sé, cioè sapere trovare l'equilibrio dinamico e vivente che riesce ad armonizzare essenza ed esistenza.

A questo riguardo Agamben, con una analisi sui concetti di *potenza destituente* e *inoperosità*, mette in luce «la capacità di disattivare e rendere inoperante qualcosa – un potere, una funzione e un'operazione umana – senza semplicemente distruggerlo, ma liberando le potenzialità che in esso erano rimaste inattuato per permetterne così un uso diverso»³⁰. L'obiettivo di Agamben, così come per Guardini, è quello di riuscire a definire un modo per salvaguardarsi dal rischio dell'onnipotenza: un rischio al quale l'umanità è stata sempre esposta, ora come in passato (come narrato dal mito di Prometeo), ma che oggi sembra portare l'umanità al crollo. Magatti, nella sua vasta analisi sul legame tra potenza, azione e relazione, mette in luce come il problema dei nostri giorni sia quello di trovare il modo di disinnescare gli aspetti distruttivi

²⁸ Cf. *Ibid.*, 236-244.

²⁹ *Ibid.*, 239.

³⁰ G. AGAMBEN, *L'uso dei corpi. Homo sacer*, IV, 2, Neri Pozza, Vicenza 2014, 345.

di una potenza che si traduce in potere incontrollabile. Il modo di neutralizzare questi effetti sta nel legame a doppio filo fra l'individuo e l'altro-da-sé:

«per un verso, l'azione è infatti la massima espressione dell'intimità personale, luogo dell'esplosione della novità di cui è portatrice; per l'altro verso, nel momento in cui si realizza, essa è già consegnata all'altro da sé, e con esso all'organizzazione sociale, che mentre la rende possibile, la limita e la definisce. Per questo, la potenza non è mai qualcosa che pertiene solo all'agire individuale. Essa, piuttosto, poggia su e si misura con le condizioni circostanti che la precedono e la seguono. Un doppio legame – che Romano Guardini chiama “polarità” – tra azione e struttura, tra forma e vita, che è essenziale per cogliere la natura insieme individuale e sociale della potenza»³¹.

Alla luce di questa riflessione, i due principi sulla *conoscenza di sé* e la *cura di sé* possono essere rilette coniugandoli al plurale. L'attuazione di essi va intesa non nella direzione di una chiusura su di sé, ma di una apertura verso l'altro-da-sé, per portare il soggetto oltre il negativo della rinuncia e farlo dirigere, in positivo, verso il guadagno di un nuovo modo di vedere il mondo e agire in esso. In un contesto fortemente mediato dalle tecnologie, fra autonomia individuale e connessione globale, questa apertura diviene condizione indispensabile per uno sviluppo equilibrato di un sé e di un *noi* all'altezza della complessità del mondo che ci stiamo costruendo attorno.

Le possibilità aperte dalle IA, dalla virtualizzazione e dalla trasposizione, in tutto o in parte, della nostra vita nell'*Onlife*, permettono alla formazione della nostra identità personale di trovare nuove vie di attuazione, ma occorre necessariamente sapersi sottrarre al gioco dell'onnipotenza, valutando le ragioni più alte del nostro sé e accettando la condizione, tipicamente umana, del limite. Agire in senso umano significa, in conclusione, avere il coraggio di attraversare – o anche bucare, all'occorrenza – il proprio guscio tecnologico per attuare una rinnovata capacità di rispondere collettivamente alle sfide globali.

³¹ MAGATTI, *Oltre l'infinito*, cit., 14.

INTELLIGENZA ARTIFICIALE E LIBERTÀ. COME CONTRASTARE LA SFIDA TRANSUMANA

Markus Krienke

Abstract

*Monitory democracy e capitalismo della sorveglianza sono due concetti sempre più utilizzati per esprimere i rischi della trasformazione digitale della società per la libertà nella società civile. Contro le sempre più analizzate strumentalizzazioni della persona a fini politici (potere) o economici (profitto), la spesso invocata esigenza di un *umanesimo digitale* può essere attuata attraverso l'idea di un *nuovo contratto sociale* che partendo dalla garanzia della *privacy* sappia fare i conti con la *vulnerabilità cognitiva* dell'essere umano nell'*infosfera*.*

Parole chiave: Transumanesimo, Capitalismo della sorveglianza, Democrazia, Dignità umana, Umanesimo digitale.

Monitory democracy and surveillance capitalism are two concepts increasingly used to express the risks of the digital transformation of society for freedom in civil society. Against the increasingly analysed instrumentalisation of the individual for political (power) or economic (profit) purposes, the often-invoked need for a digital humanism can be realised through the idea of a new social contract that, starting from the guarantee of privacy, is able to come to terms with the cognitive vulnerability of the human being in the infosphere.

Keywords: Transhumanism, Surveillance capitalism, Democracy, Human dignity, Digital humanism.

«Non sarà un demone a scegliere voi,
ma sarete voi a scegliere il vostro demone».

Platone, *Repubblica X*, 617e

«L'inizio e la fine di ogni filosofia è – libertà!»

Friedrich Wilhelm Joseph Schelling,
Dell'io come principio della filosofia, § 6

1. Introduzione: società transumana e umanità "antiquata"

Sebbene intorno alle nuove tecnologie tra IA e ChatGPT si riconfigurino tutte le categorie del "politico" – dalla sovranità al potere, dalla democrazia alla partecipazione, dalla giustizia sociale alla garanzia dei diritti fondamentali – non saranno certamente le tecnologie digitali stesse (in forma di intelligenza artificiale generale (GAI), superintelligenza o singolarità) a esercitare il potere in futuro¹: «[n]essuno Spartaco artificiale guiderà una rivolta digitale»². Ma il fatto che non si tratti di tecnologie *neutre*³ – se mai di tali si può parlare in quanto persino il

¹ Certamente «questa idea di una macchina che, appena diventata più intelligente di noi, si preoccupa immediatamente di dominarci e magari perfino di annientarci» non è altro che «un possibile caso di antropomorfizzazione» (T. POGGIO - M. MAGRINI, *Cervelli Menti Algoritmi*, Sperling & Kupfer, Milano 2023, 216); cf. T. FUCHS, *Verteidigung des Menschen. Grundfragen einer verkörperten Anthropologie*, Suhrkamp, Frankfurt a. M. 2020, 65.

² L. FLORIDI, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina, Milano 2022.

³ Günther Anders sottolinea che «la credenza che esistano porzioni del nostro mondo che non sono altro che "mezzi", a cui si possono assegnare *ad libitum* "scopi buoni", sono pure illusioni. Radio e televisione stesse sono realtà, e realtà che ci plasmano. E questo dato di fatto, che ci plasmano qualunque sia lo scopo al quale le impieghiamo, non viene eliminato soltanto perché noi le degradingo verbalmente a "mezzi". [...] la vera e propria *humanitas* comincia soltanto là dove questa distinzione perde ogni senso: dove tanto i mezzi che gli scopi sono talmente impregnati di stile di vita e di etica che, al cospetto di singole porzioni di vita o di mondo, non si può distinguere, anzi non si affaccia nemmeno l'interrogativo se si tratti di "mezzi" o di "scopi" Ogni mezzo è più di un

classico “coltello” spesso richiamato in causa deve avere le caratteristiche giuste qualora volesse essere usato per “uccidere”, e viceversa si può impedire un tale utilizzo costruendolo in modo alternativo – le rende incisive per la configurazione e la comprensione dello spazio sociale. In questo senso le trasformazioni sociali sono *già in atto* in quanto il semplice utilizzo delle nuove tecnologie da parte di quasi tutta la società sta cambiando gradualmente il modo di determinare e legittimare il potere politico, la possibilità di produrre e distribuire beni e ricchezza, di lavorare e amministrare, di vivere gli spazi pubblici e privati, ecc. E questo cambiamento si realizza anche per chi non utilizza – ugualmente se non “può” o non “vuole” – queste tecnologie. Senz’altro, gli “imperativi” dell’evoluzione tecnologica, nonché la preoccupazione di “non perdere il treno”, da un lato, e le paure di grandi parti della società insieme alla richiesta di un aumento della regolamentazione e dell’etica in generale nel settore, dall’altro, distolgono lo sguardo da una chiara percezione di tali dinamiche trasformative. In questo senso, la tecnica digitale non è più “mezzo” o “strumento” per la soluzione di problemi, ma è *design* sociale che «ridisegna [...] la geografia del potere e il suo sistema di *checks and balances*»⁴. In qualche modo la caratteristica della tecnica come *design* riprende l’interpretazione heideggeriana che l’identificava con l’«imposizione» (*Gestell*)⁵, e realizza ciò che Günther

mezzo. Gli strumenti stessi non sono dei semplici oggetti che si possono utilizzare ma determinano già, con la loro struttura e la loro funzione, il loro utilizzo e lo stile delle nostre attività e della nostra vita, in breve, ci determinano», G. ANDERS, *L’uomo è antiquato*, vol. 1: *Considerazioni sull’anima nell’epoca della seconda rivoluzione industriale*, Bollati Boringhieri, Torino 2007, 97-98.

⁴ P. STANZIONE, *Intelligenza artificiale e decisioni politiche*, in A. PATRONI GRIFFI (cur.), *Bioetica, Diritti e intelligenza artificiale*, Mimesis, Milano-Udine 2023, 215-224, qui 216.

⁵ «L’essenza della tecnica moderna consiste nell’imposizione [*Gestell*]. Questa appartiene al destino del disvelamento. [...] Il destino del disvelamento è in se stesso non un pericolo qualunque, ma *il pericolo*», M. HEIDEGGER, *La questione della tecnica*, in *Id.*, *Saggi e discorsi*, a cura di G. Vattimo, Mursia, Milano 1991, 5-27, qui 19-20.

Anders intendeva quando la definiva il «soggetto della storia»⁶.

In quanto questi processi sociali, e ancora più in profondità l'antropologia che sta alla loro base (secondo Han risultiamo «riprogrammati»⁷), subiscono delle trasformazioni paradigmatiche tramite l'evoluzione e l'utilizzo delle nuove tecnologie, tali trasformazioni si ripercuotono anche sulla sfera politica ed economica, costituendo una sfida non trascurabile alla libertà. Nella misura in cui le nostre libertà, con il crescente utilizzo delle nuove tecnologie, stanno diminuendo, il *transumanesimo* è già in atto, e questa interpretazione certamente dà a tale termine un'accezione differente da quella con cui viene utilizzato dai suoi fautori⁸.

Per il transumanista Yuval Noah Harari, il *dataismo* (o «datismo») esprime uno scetticismo «riguardo alla conoscenza e alla saggezza umane, e preferisc[e] riporre la [...] fiducia nei Big Data e negli algoritmi computerizzati»⁹, riconfigurando quel meccanismo che, secondo Habermas, è decisivo per la formazione della *sfera pubblica* nella modernità, ossia una sfera di intermediazione tra società e politica istituzionalizzata, dove le interazioni sociali si riflettono in modo razionale e critico, in cui vale il principio della trasparenza razionale cioè della precedenza degli argomenti migliori contro una politica dell'arbitrio¹⁰. Lo stesso

⁶ G. ANDERS, *L'uomo è antiquato*, vol. 2: *Sulla distruzione della vita nell'epoca della terza rivoluzione industriale*, Bollati Boringhieri, Torino 1992, 258 (corsivo nell'originale).

⁷ B.C. HAN, *Nello sciame. Visioni del digitale*, Nottetempo, Roma 2015, 9.

⁸ Per loro transumanesimo significa «un movimento culturale e intellettuale che afferma che è possibile e desiderabile migliorare fundamentalmente la condizione umana mediante l'uso della ragione, in particolare sviluppando e diffondendo ampiamente le tecniche volte a eliminare l'invecchiamento e a migliorare in modo significativo le capacità intellettuali, fisiche e psicologiche dell'essere umano», *The Transhumanist Declaration*, cit. in: D. FOLSCHIED - A. LÉCU - B. DE MALHERBE, *Che cos'è il transumanesimo?*, Queriniana, Brescia 2021, 15.

⁹ Y. N. HARARI, *Homo deus. Breve storia del futuro*, Bompiani, Milano 2020, 450.

¹⁰ Egli sottolinea la necessità che questa «formazione dell'opinione e della volontà», istituzionalizzata nella democrazia, resti *autonoma* dal meccanismo politico: essa avviene grazie agli «spontanei, non manipolati flussi comunicativi di una sfera pubblica non programmata per decidere», J. HABERMAS, *Morale, Diritto, Politica*, a cura di L. Ceppa, Einaudi, Torino 2007, 95-96.

Habermas vede questi processi, essenziali per la democrazia, messi a rischio dalla trasformazione digitale della società che depolitizza la sfera pubblica rendendola uno spazio «di intrattenimento e di consumo»¹¹. Secondo Han i «dataisti contrappongono alla teoria habermasiana dell'agire comunicativo una *teoria behaviorista dell'informazione*, che procede senza discorso. Nella visione dataista del mondo non c'è spazio per l'individuo che agisce razionalmente, che avanza una pretesa di legittimità e la difende argomentando»¹². Sempre Han analizza come il modello behavioristico interpreti – e poi indirizzi e diriga – la società senza ricorrere ai presupposti della libertà e autodeterminazione della persona. Con i dati si ottimizzano le prestazioni sia della democrazia sia del capitalismo, come hanno analizzato rispettivamente Luciano Floridi e Shoshana Zuboff: populismo e capitalismo della sorveglianza sono i nomi che indicano la perdita della libertà nell'ambito politico ed economico, nella misura in cui con le dinamiche transumane del dataismo il soggetto viene «cancellato, come sull'orlo del mare un volto di sabbia»¹³. Per dirlo ancora una volta: non è il *problema tecnologico* dell'evoluzione dell'intelligenza artificiale e delle tecnologie digitali ciò che costituisce la vera sfida e rende il *transumanesimo* l'espressione della preoccupazione per la libertà e la dignità umana come criteri centrali dei sistemi economici e politici, ma il fatto che potenti centri di potere (di carattere economico, finanziario, politico, ecc.) si impadroniscono delle sue possibilità per pianificare i processi sociali e le loro garanzie istituzionali. Per questo risulta corretto affermare che «[l]a concentrazione di potenza digitale oggi è una minaccia potenziale per la democrazia

¹¹ J. HABERMAS, *Nuovo mutamento della sfera pubblica e politica deliberativa*, Raffaello Cortina, Milano 2023, 60.

¹² B.C. HAN, *Infocrazia. Le nostre vite manipolate dalla rete*, Einaudi, Torino 2023, 50.

¹³ M. FOUCAULT, *Les Mots et les Choses. Une archéologie des sciences humaines*, Gallimard, Paris 1966, 398 (tutte le traduzioni in questo contributo, se non indicato diversamente, sono mie).

e per i mercati»¹⁴. Viceversa, la dinamica transumanista significa che «[p]er la società strumentalizzata l'individualità è una minaccia, un ostacolo problematico che toglie energie a "collaborazione", "armonia" e "integrazione"»¹⁵. Di fronte all'*enhancement* dell'ottimizzazione dei processi politici ed economici, la libertà d'arbitrio non è altro che un residuo illusorio di una "modernità antiquata"¹⁶.

2. Che cosa avviene nell'infosfera?

Tirando le somme da queste considerazioni introduttive, si può affermare con Gigerenzer che «[l]a tecnologia non è solo un sistema di supporto; richiede a noi di adattare il nostro comportamento»¹⁷. Solo un tale adattamento permette alle nuove tecnologie di funzionare al meglio; ma tale funzionamento consiste anche nell'«usare in modo strumentale ipotesi anticipatrici di futuro per promuovere e indirizzare l'opinione pubblica contemporanea»¹⁸. In questo modo l'informazione diventa la categoria universale interpretativa e predittiva della realtà sociale¹⁹. Secondo l'interpretazione radicale di Hofstetter, «la sorveglianza si è imposta come

¹⁴ I. IACONE, *Democrazia precaria. Dominio degli algoritmi e condizione umana*, in *Prospettiva Persona* 31 (2023) 2, 58-70, qui 62.

¹⁵ S. ZUBOFF, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Luiss University Press, Roma 2019, 454.

¹⁶ Scrisse Anders già nel 1956 che a causa «del "dislivello prometeico" la nostra propria metamorfosi è in ritardo; la nostra anima è rimasta molto indietro in confronto al punto a cui è arrivata la metamorfosi dei nostri prodotti, ossia del nostro mondo», ANDERS, *L'uomo è antiquato*, I, cit., 26.

¹⁷ G. GIGERENZER, *Perché l'intelligenza umana batte ancora gli algoritmi*, Raffaello Cortina, Milano 2023, 94.

¹⁸ L. GRION, *Chi ha paura del post-umano? Vademecum dell'uomo 2.0*, Mimesis, Milano-Udine 2021, 94.

¹⁹ Sottolinea infatti Mainzer che «il concetto di informazione è la categoria universale con cui si possono catturare non solo i cambiamenti tecnici, ma anche quelli sociali, economici e societari», K. MAINZER, *Künstliche Intelligenz - Wann übernehmen die Maschinen*, Springer, Berlin 2019, 230.

la nuova normalità. Gafam [= acronimo che comprende le cinque maggiori multinazionali della tecnologia: Google, Apple, Facebook, Amazon, Microsoft] non hanno soltanto potere economico, ma sono loro i padroni nel mondo digitale, non l'apparato classico dello Stato»²⁰.

Questo significa che cambia fundamentalmente ciò che le tecnologie "fanno": non si tratta più di "spiegare" ma di "predire". I «pattern dovrebbero cioè consentire di prevedere gli sviluppi futuri, anche se non necessariamente li si può spiegare. La modellazione predittiva è diversa dalla modellazione esplicativa»²¹. La comprensione avviene sempre in un momento secondario rispetto alla previsione, che però non mira a "vedere il futuro" come nelle tecniche antiche divinatorie, bensì a "pianificare", cioè ad adattare il presente a un futuro che ovviamente rimane in sé sconosciuto²²: in altre parole, gli algoritmi "plasma-no" il futuro modificando il presente. In questo modo, l'algoritmo in un certo senso "ha sempre ragione", in quanto viene superato il classico concetto di "verità".

Come ripercussione sociale di tale dinamica adattiva, ne risulta una società progredita nel suo processo di individualizzazione ed articolazione del narcisismo: «[a]llo sciame digitale manca l'anima della folla o lo spirito della folla: gli individui che si uniscono in uno sciame non sviluppano un *Noi*»²³. Con ciò, Han descrive la caratteristica dinamica

²⁰ Y. HOFSTETTER, *Das Ende der Demokratie. Wie die künstliche Intelligenz die Politik übernimmt und uns entmündigt*, Penguin, München 2018, 371.

²¹ E. ESPOSITO, *Comunicazione artificiale. Come gli algoritmi producono intelligenza sociale*, Bocconi University Press, Milano 2022, 146-147: «I termini utilizzati nella previsione algoritmica ("correlazioni", "pattern"), l'idea di previsioni indipendenti da relazioni causali, il riferimento a strutture inaccessibili al ragionamento umano, hanno tutti una tradizione antica e complessa nelle società divinatorie [...]. In entrambi i casi si tratta di programmi che, a differenza delle pratiche scientifiche, non cercano di spiegare o di comprendere i fenomeni, ma solo di gestirli».

²² La differenza con la statistica sta nel fatto che «[l']obiettivo della statistica è gestire l'incertezza presente. Si occupa della conoscenza (o della mancanza di conoscenza) del presente, mantenendo e confermando l'insuperabile barriera tra il momento attuale e il futuro aperto. Il *machine learning*, invece, si rivolge al futuro e cerca di prevederlo», *Ibid.*

²³ HAN, *Nello sciame*, cit., 22.

dell'isolamento che si realizza nell'infosfera. Proprio per questo gli individui contemporanei sono incapaci di formare un peso politico se per esso è essenziale la continuità di presenza e la realizzazione di una forza capace di incidere sulla realtà. Questa dinamica ha luogo perché alla base viene sottratto alla società il concetto di "verità", vale a dire di riferirsi insieme, attraverso un "noi", a una realtà comune e solida. «Le fake news non sono menzogne: esse attaccano la fatticità stessa. De-fatticizzano la realtà. [...] La crisi della verità mina la fiducia nei fatti stessi»²⁴. Alle *grandi narrazioni* che, secondo Lyotard, nel mondo post-moderno sono ormai "terminate", si sostituiscono le «*micro narrazioni delle teorie del complotto*», che nel mondo digitale «sono assunte come *risorse identitarie e di significato*»²⁵. Di fronte a tale scenario, non è certamente una re-ideologizzazione della società – nel senso di programmi politici "di destra" o "di sinistra"²⁶ – a costituire la risposta adeguata alla forza della trasformazione in atto. Prima di avanzare qualche ragionamento a riguardo, occorre analizzare più concretamente le ripercussioni di tale trasformazione sulle dinamiche del libero mercato e della democrazia.

Innanzitutto, la trasformazione sociale causa una trasmutazione dell'economia libera, ormai globalizzata, in un *capitalismo della sorveglianza*. Come all'inizio della modernità, e cioè nella prima «Grande Trasformazione» (Polanyi), l'introduzione del principio della *divisione del lavoro*, oltre a cambiare il processo industriale di produzione, ridisegnava la società in quanto tale, così nella seconda Grande Trasformazione per «noi, membri della seconda modernità, la divisione dell'apprendimento è quello che la divisione del lavoro era per i nostri nonni e bisnonni», e questa viene controllata «da parte del capitalismo della sorveglianza»²⁷, il quale, tramite tutte le informazioni che emettiamo nel mondo digitale («primo testo»), produce un «secondo testo» o «testo

²⁴ *Ibid.*, 62-63.

²⁵ *Ibid.*, 72.

²⁶ Cf. V. COSTA, *Categorie della politica. Dopo destra e sinistra*, Rogas, Roma 2023.

²⁷ ZUBOFF, *Il capitalismo*, cit., 198-199.

ombra» che rappresenta il «surplus» consistente nella «nostra esperienza [che] viene costretta a diventare una materia prima da accumulare e analizzare per i fini commerciali di altre persone»²⁸. Tale sapere viene sfruttato unilateralmente a scopi economici, introducendo un enorme divario sociale nella società, capace di nuocere sensibilmente il tessuto integrativo della democrazia. «In questo movimento nascosto, la competizione per i guadagni della sorveglianza ci riporta a un ordine pre-Gutenberg, con la divisione dell'apprendimento nella società che diviene patologica, riservandolo a uno stretto clero di specialisti informatici che lavorano per privati con macchine private che imparano per i loro interessi economici»²⁹. Le tecnologie consentono, quindi, ingenti accumulazioni di ricchezza nelle mani delle più grandi imprese dell'IA, che non soltanto ottengono una posizione economica concorrenziale a molti Stati, ma garantiscono una infrastruttura ormai indispensabile: ma allora vale anche per queste grandi imprese – come per le grandi banche – l'essere *too big to fail*, con tutte le conseguenze sul piano degli impegni politici?

Il capitalismo della sorveglianza inizia ovunque si paga non con denaro, ma con i dati personali per un determinato servizio³⁰, in quanto questi sono la base per la produzione relativamente affidabile di comportamenti che poi effettuano risultati economicamente rilevanti: la «modifica automatizzata del comportamento» è il suo scopo principale³¹. Perché è chiaro che più dati si riescono a raccogliere e più il *microtargeting* si perfeziona. A questa strategia economica si associa quella politica del *big nudging*, che è un altro utilizzo dei dati con l'intenzione di far comportare il cittadino in una determinata maniera, pubblicamente "bravo", ma anche per esprimere una determinata preferenza di voto. Perciò, secondo Benanti e Maffettone, «una visione libertaria dell'universo digitale», come era possibile sperare all'inizio

²⁸ *Ibid.*, 199.

²⁹ *Ibid.*, 203.

³⁰ Cf. GIGERENZER, *Perché l'intelligenza umana*, cit., 215-293.

³¹ ZUBOFF, *Il capitalismo*, cit., 313.

della trasformazione digitale, allo stato attuale dell'evoluzione non è più concepibile³². Tutto questo ha delle ripercussioni sulla democrazia e il modo in cui si articola nell'infosfera.

La sostituzione di azioni libere con il rilevamento di dati circa il comportamento degli individui avviene anche a livello politico, dove si realizzano forme di post-democrazia. Già i *bias* esprimono una sfida per la democrazia e l'etica, e secondo Hofstetter «[i]ncombe l'interdizione del cittadino. C'è la minaccia del dominio di un'élite che dispone dei dati, dei mezzi di elaborarli e del capitale – ugualmente su tutti i mezzi di produzione – del capitalismo dell'informazione»³³. Questa minaccia viene alimentata da particolari patologie che si articolano grazie alle possibilità tecnologiche dei nuovi media «come hate speech e forme di polarizzazione nella comunicazione pubblica»³⁴, ma si realizza soprattutto tramite la decostruzione della *sfera pubblica* come effetto della trasformazione digitale della società. Mentre le concrete tecniche di manipolazione dell'opinione pubblica rappresentano un rischio per la democrazia, ma hanno effetti solo a breve termine – gli stessi movimenti di antipolitica o i partiti populistici “di piattaforma” non riescono a trasformare il campo dei partiti in maniera sostanziale – sono le erosioni delle infrastrutture della democrazia stessa il vero potenziale di rischio a medio e lungo termine di queste trasformazioni. Lo stesso Habermas esprime grande preoccupazione per la trasmutazione della sfera pubblica da un libero scambio di opinioni, di cui l'indipendenza del giornalismo aveva assicurato un certo livello di razionalità e neutralità delle informazioni, alla gestione delle informazioni sulle piattaforme digitali

³² P. BENANTI - S. MAFFETTONI, *Decisioni politiche e sostenibilità digitale*, in P. SEVERINO (cur.), *Intelligenza artificiale. Politica, economica, diritto, tecnologia*, Luiss University Press, Roma 2022, 9-30, qui 21.

³³ HOFSTETTER, *Das Ende der Demokratie*, cit., 350. L'autrice giunge al risultato che le «nuove tecnologie digitali e i modelli di business sono un rischio per la società democratica. Come capitalismo dell'informazione disattendono i diritti fondamentali di una società democratica» (*ibid.*, 402).

³⁴ R. FISCHER - O. JARREN, *The platformization of the public sphere and its challenge to democracy*, in *Philosophy and Social Criticism* 50 (2024), 1, 200-215, 16, qui 211.

tramite algoritmi sviluppati da aziende private³⁵. La frammentazione della società civile causata dai *digital media* costituisce una minaccia alla democrazia in quanto favorisce la "personalizzazione" della comunicazione politica – specialmente dei partiti e dei loro *leader* – a scapito della formazione di un consenso democratico nella sfera pubblica che presuppone il rapporto di tutti i cittadini a una comune realtà condivisa³⁶. Al capitalismo della sorveglianza corrisponde quindi, in questa visione "transumana", il rischio di finire in una *monitory democracy*³⁷. Concretamente, con le tecnologie del *Demos scraping* le votazioni in democrazia vengono sostituite dal rilevamento di preferenze e comportamenti dei cittadini, che consentono di produrre risultati molto più esatti rispetto ai classici sondaggi. Le politiche, in questo scenario, non sono la realizzazione partecipativa di cittadini autodeterminati, ma "soluzioni" adatte a una situazione mappata di opinioni e preferenze da parte dei cittadini e dunque più capaci, rispetto ai tempi passati, di incontrare e accedere alla volontà del popolo³⁸.

Già Olivetti riteneva in crisi «la democrazia e il parlamentarismo», mentre «la rappresentanza tradizionale e i parlamenti eletti dal suffragio

³⁵ «La portata dell'articolazione dei media indipendenti deve essere abbastanza forte, in modo da non spezzare il legame esistente tra il potere politico e il potere comunicativo generato dai cittadini – che è l'unica "autorità" che è "emanata" dal popolo», HABERMAS, *Nuovo mutamento*, cit., 112.

³⁶ «Secondo la teoria democratica deliberativa, la democrazia contemporanea dipende dalla sfera pubblica politica come luogo non solo di libertà, ma anche di razionalità [...]. È quindi necessario creare uno spazio di razionalità socialmente, culturalmente e politicamente condiviso che non sia prevalentemente determinato dalla logica economica», C. RITZI, *Politische Öffentlichkeit zwischen Vielfalt und Fragmentierung*, in J. HOFMANN et al. (hrsg.), *Politik in der digitalen Gesellschaft. Zentrale Problemfelder und Forschungsperspektiven*, transcript Verlag, Bielefeld 2019, qui p. 75.

³⁷ J. KEANE, *Democracy and media decadence*, Cambridge University Press, Cambridge-New York 2013, 79.

³⁸ Cf. L. ULBRICHT, *Scraping the demos. Digitalization, web scraping and the democratic project*, in *Democratization* 27 (2020) 3, 426-442, qui 432. Secondo l'autrice, il «demos scraping, nella sua forma attuale, è un cavallo di Troia per il capitalismo di sorveglianza tecnocratico e per la materializzazione anestetica della democrazia simulativa» (*ibid.*, 435).

universale non funzionano più». Si sarebbe giunti al momento in cui il «compito dei partiti politici sarà esaurito e la politica avrà un fine quando sarà annullata la distanza fra mezzi e fini, quando cioè la struttura dello stato e della società giungeranno a un'integrazione, a un equilibrio, per cui sarà la società e non i partiti a creare lo stato»³⁹. E così avvenne, senza però produrre l'auspicato «equilibrio»: è la *crisi* della democrazia, che si configura attualmente come la crisi dei partiti in quanto questi ultimi vedono, ormai, la loro funzione più specifica – la formulazione, propaganda e realizzazione politica di *programmi* – realizzata da algoritmi⁴⁰.

3. Un «nuovo contratto sociale»⁴¹?

Non è soltanto la disintermediazione⁴², e quindi la sottrazione dell'indipendenza della sfera civile alla pianificazione politica⁴³, ciò che si

³⁹ A. OLIVETTI, *Democrazia senza partiti. Fini e fine della politica*, Edizioni di Comunità, Roma-Ivrea 2013, 28-29.

⁴⁰ Per questo, Garasic chiama la «dipendenza e sottomissione che i programmi politici hanno nei confronti degli algoritmi» con il termine «algorpolitica», M.D. GARASIC, *Leviatano 4.0. Politica delle nuove tecnologie*, Luiss University Press, Roma 2022, 40. Concretamente, egli rileva che «la scelta di esprimersi in 280 caratteri, pensare attraverso hashtags invece che programmi elettorali e politici, generare visibilità attraverso una spettacolarizzazione (in realtà antecedente all'era digitale) del motto, della frase, dell'invettiva sono tutte azioni utili ad apparire e a finire velocemente nel dimenticatoio. In altre parole la versione più distopica della precedentemente citata politica quantistica» (*ibid.*, 49). Se la situazione però è questa, non si capisce come – ed è quello l'altro e più "positivo" significato del termine per Garasic – queste tecnologie debbano allo stesso momento mettere la politica «in grado di sviluppare anticorpi utili a non far inghiottire la politica dagli algoritmi» (*ibid.*).

⁴¹ GRUPPO DI ESPERTI MISE SULL'INTELLIGENZA ARTIFICIALE, *Proposte per una strategia italiana per l'intelligenza artificiale*, luglio 2019, www.mise.gov.it (accesso 1.10.2023).

⁴² Nell'infosfera, «[m]ediazione e rappresentazione vengono interpretate come mancanza di trasparenza e inefficienza, come un ristagno di tempo e informazioni», HAN, *Nello sciamme*, cit., 29.

⁴³ «Quanto qualcosa del genere sia pervasivo e se le conseguenze vadano in direzione di maggiore democraticità o piuttosto verso un rischio di nuovi totalitarismi della rete, è difficile dire», BENANTI - MAFFETTONE, *Decisioni politiche*, cit., 21.

realizza nel mondo dei nuovi *media*. Si può osservare che essa viene controbilanciata da «una inedita forma di intermediazione, che è diversa da quella praticata dai media cosiddetti tradizionali. Essa avviene attraverso l'algoritmo, ovvero un certo modo di governare il flusso delle informazioni e quindi costruire le narrazioni»⁴⁴. Se, infatti, dovesse avere alcuna plausibilità la tesi che attraverso una *distruzione creativa* – quale caratteristica di ogni innovazione tecnologica – la “società analogica” si trasforma in uno spazio d'azione sociale diverso, allora sarà possibile pensare non solo la riaffermazione della democrazia – in nuove forme partecipative – nell'*infosfera* o *digital matrix*, ma addirittura l'apertura di nuove possibilità in cui essa riesce a realizzarsi. In questo senso, contro le analisi negative di Habermas, che riprende uno scetticismo critico di Arendt e si conferma nelle considerazioni di Honneth, Capurro sottolinea che «Internet consente l'accesso all'informazione soprattutto ai gruppi emarginati, ma anche perché questi ultimi possono formare le proprie reti»⁴⁵.

Al fine di comprendere queste nuove forme di espressione della libertà democratica che aprono lo spazio a un «umanesimo digitale»⁴⁶, Jeanette Hofmann è contraria a considerare il rapporto tra «democrazia» e «digitalizzazione» secondo un semplice schema di «causalità» tra entrambe, perché “la digitalizzazione” non è un soggetto agente, e tanto meno un'entità statica, laddove nella sua evoluzione è sempre *embedded* in dinamiche sociali con le quali sta in un rapporto di reciproco condizionamento⁴⁷. In tale prospettiva si evince immediatamente come

⁴⁴ C. DANANI, *Abitare il futuro tecnologico: quale dimensione politica*, in L. SANDONÀ-F. ZACCARON (curr.), *Noi techno-mani. Pandemia, tecnologie, comunità*, in *Anthropologica. Annuario di studi filosofici* (2020) 137-147, qui 139-140.

⁴⁵ R. CAPURRO, *Homo Digitalis. Beiträge zur Ontologie, Anthropologie und Ethik der digitalen Technik*, Springer, Wiesbaden 2017, 179.

⁴⁶ J. NIDA-RÜMELIN - N. WEIDENFELD, *Umanesimo digitale. Un'etica per l'epoca dell'Intelligenza Artificiale*, Franco Angeli, Milano 2019.

⁴⁷ Cf. J. HOFMANN, *Mediatizzata Demokratie in Zeiten der Digitalisierung. Eine Forschungsperspektive*, in EAD. et al. (hrsg.), *Politik in der digitalen Gesellschaft*, cit., 27-45, qui 28.

alla maggior parte delle analisi di questo rapporto resti inosservato che è «il cambiamento delle istituzioni e delle pratiche della democrazia rappresentativa ciò che effettivamente rende plausibile la forma e l'uso attuale dei media digitali»⁴⁸. Solo in questo modo si prende sul serio il risultato dell'antropologia filosofica novecentesca, che ha evidenziato come la tecnica sia parte della *conditio humana* e strutturi pertanto il campo di possibilità per l'agire umano. Questa specifica forma l'*internet* l'ha assunta come «modello opposto a una rete statale di comunicazione»: rispetto a quest'ultima tecnologia a disposizione negli anni '90 si è affermata vincente l'odierna forma di tecnologia rispetto alla quale esistevano delle alternative⁴⁹. Se si considera la trasformazione digitale in questo modo, ossia come *cambiamento di forma* della democrazia in direzione «post-rappresentativa»⁵⁰, rivolta a un pubblico non più passivo ma attivo, non più al «voter citizen» ma al «vigilant citizen», allora diventa chiaro che una politica incentrata sull'organizzazione partitica deve fare spazio a una orientata ai temi⁵¹. In questa prospettiva, *liking*, *following* e *(re)tweeting* sono le nuove forme in cui si esprime la partecipazione nella nostra tarda modernità. Secondo Hofmann, queste evoluzioni non sono una rinuncia dei cittadini alla politica rappresentativa, ma piuttosto la loro *reazione* all'indebolirsi di quest'ultima⁵².

Contrariamente ad autori come Capurro e Hofmann, Luciano Floridi analizza l'interpretazione della politica da parte dei suoi principali attori, specialmente politici e partiti, osservando la tendenza a una

⁴⁸ *Ibid.*, 29.

⁴⁹ *Ibid.*, 35.

⁵⁰ *Ibid.*, 37-38.

⁵¹ Cf. *ibid.*, 37 e 39.

⁵² «Nel contesto di questo cambiamento di significato, si sta verificando la trasformazione di un pubblico di destinatari, un tempo passivo, in una società civile che interviene discorsivamente. A mio avviso, attribuire questo sviluppo esclusivamente o principalmente ad altre tecnologie digitali, come fanno molti osservatori, sopravvaluta l'agenzia dei media e allo stesso tempo sottovaluta la spinta politica derivante dal declino della capacità rappresentativa del panorama partitico e dei suoi mezzi di comunicazione di massa», *ibid.*, 41, traduzione mia.

«marketizzazione della politica» con la dinamica “transumana”, in cui il «marketing vede e usa le persone come interfacce»⁵³. Dato che le nuove tecnologie hanno trasformato le *persone* in *interfacce*, cioè «spazi di interazioni tra, da un lato, *agenti* [...] che vogliono qualcosa da noi e, dall’altro, qualcosa che [...] abbiamo, e che questi agenti desiderano o vogliono, le nostre *risorse*»⁵⁴ (in questo caso il consenso in forma di voto) – la politica non si orienta più alla questione della “verità”, ma al criterio se “funziona” e quindi se “ha successo”. Proprio per questo, conclude Floridi, non bisogna cadere nell’illusione che il principio rappresentativo della democrazia si lasci sostituire dalla tecnologia digitale, pensando di realizzare una democrazia che sia “migliore” soltanto perché “più diretta”⁵⁵. In questa direzione, Zuboff critica il presupposto dell’inevitabilità utopistica della realizzazione della trasformazione tecnologica della società, che opera perfettamente secondo uno schema marxista, rendendo gli individui passivi nei confronti di una logica di sviluppo materialistica. Dunque, niente nuoce alla democrazia quanto l’illusione di poterla perfezionare tramite le nuove tecnologie⁵⁶: «[s]e

⁵³ L. FLORIDI, *Il verde e il blu. Idee ingenue per migliorare la politica*, Raffaello Cortina, Milano 2020, 186 e 183.

⁵⁴ *Ibid.*, 138. «Per il mondo dei *social media*, siamo interfacce tra essi e i nostri *dati personali*. Per il mondo del *commercio*, siamo interfacce tra esso e il nostro *denaro* o *credito*. E per il mondo della *politica*, siamo interfacce tra essa e la nostra *attenzione*, i nostri *consenso* e *voto*», *ibid.*, 139.

⁵⁵ «La terapia non è più democrazia, che aumenta il populismo, o meno democrazia, che aumenta l’illiberalismo, ma una democrazia migliore, che sia antipopulista, liberale e politica, per contrastare la dittatura della maggioranza e rappresentare, difendere e riconciliare gli interessi legittimi di tutti, anche le minoranze», *ibid.*, 221-222.

⁵⁶ Una tale posizione si esprime ad esempio nella seguente affermazione di Honkela: «Vorrei suggerire che la democrazia può essere rafforzata in modo che tutte le persone possano partecipare al processo decisionale. Non dobbiamo scegliere uno o pochi leader, ma possiamo coinvolgere tutto il popolo, se lo desideriamo» (*Timo Honkela, and his keynote speech on the second day of April's National Dialogues conference is titled Peace from a Different Perspective – a Dialogue of a Million People*, in: <https://www.aljazeera.com/features/2017/5/30/could-artificial-intelligence-lead-to-world-peace> (accesso 7.10.2023). Tale idea era stata avanzata anche da B.C. HAN nel suo saggio *Razionalità digitale. La fine dell’agire comunicativo*, GoWare, Firenze 2014.

tutti sapessero tutto, la società sarebbe impossibile»⁵⁷. Pertanto la domanda che molti si pongono, ovvero se sia pensabile che in un futuro non lontano ci si possa abituare alle probabilità tecnologiche, finendo con preferire quest'ultime alle regole generali prodotte dalla democrazia, svela una prospettiva preoccupante.

In queste dinamiche (siano esse interpretate in senso positivo o negativo per la democrazia) molti interpreti vedono erodere il fondamento della società moderna stessa, il *contratto sociale*: nella misura in cui svanisce la realtà della persona, mancano i soggetti di questo contratto; in altre parole, è questo contratto, quale fondamento della società nell'epoca moderna, a non essere più adatto a esprimere la garanzia per la libertà e autodeterminazione dei soggetti. Infatti Habermas osserva, con preoccupazione, che «[i]n determinate subculture, la sfera pubblica non è più percepita come inclusiva, così come la sfera pubblica politica non è più percepita come uno spazio di comunicazione volto alla generalizzazione degli interessi comprendente tutti i cittadini»⁵⁸. Bisogna ripensare la sfera civile in maniera opposta alla riduzione del cittadino a *interfaccia*. Alcuni approcci realizzano questa esigenza a partire dalla pretesa di una maggiore garanzia della *privacy*, da interpretare come «un *habeas data*: corrispettivo, nella società digitale, di ciò che l'*habeas corpus* ha rappresentato sin dalla *Magna Charta*»⁵⁹, vale a dire l'inizio e la base di ogni contratto sociale, e ciò significa anche il divieto di tecnologie capaci di anticipare in maniera sensibile il giudizio degli individui, intervenendo sulla loro capacità di autodeterminazione (ecco l'ambito dei "neurodiritti"). La protezione della *privacy* segna, quindi, una delle «nuove vulnerabilità al danno» nell'*infosfera*⁶⁰. La *privacy*, in

⁵⁷ S. SAMPSON, *The Morality of Transparency: Clarity versus Emptiness*, in V. AUGUST - F. OSRECKI (hrsgs.), *Der Transparenz Imperativ. Normen - Praktiken - Strukturen*, Springer VS, Wiesbaden 2020, 37-62, qui 40.

⁵⁸ HABERMAS, *Nuovo mutamento*, cit., 64-65.

⁵⁹ STANZIONE, *Intelligenza artificiale*, cit., 220.

⁶⁰ G. MANZONE, *Morale artificiale. Nanotecnologie, intelligenza artificiale, robot. Sfide e promesse*, EDB, Bologna 2020, 72.

questa prospettiva, è così importante perché solo essa può garantire il riconoscimento della *persona*. La persona, come ha sottolineato Spaemann, è caratterizzata dalla capacità di riflettere, cioè di acquisire una *distanza* rispetto ai propri interessi, che «cessano di essere, nella loro immediatezza, i criteri che orientano la riflessione»⁶¹. Riconoscendo in un compito politico, che miri alla ricostituzione del contratto sociale attraverso la garanzia della *privacy*, la principale strada della realizzazione di un tanto atteso *umanesimo digitale*, molti autori formulano delle richieste assai concrete, come ad esempio il riconoscimento «che la *privacy* di un individuo o di un gruppo viene violata anche quando le informazioni sensibili che lo riguardano vengono segnalate a sua insaputa e contro la sua volontà»⁶².

Questo discorso attualmente acceso circa una più efficace protezione della *privacy* pone un'ulteriore questione importante: all'interno di un nuovo *contratto sociale* dovrebbe essere ancorata anche la specifica responsabilità sociale e politica che le aziende delle nuove tecnologie realizzano di fatto, considerando che con le loro infrastrutture esercitano un compito *pubblico* che deve essere definito in quanto tale. Ora, per quanto realizzano anche questa dimensione con la loro *agency*, tali aziende sono già uscite dalla sfera *privata* e dovrebbero essere suscettibili di una legislazione⁶³.

Un ulteriore spunto riguarda l'ipotesi di ragionare sulla garanzia del «diritto al proprio tempo», sia per quanto riguarda l'accelerazione dei processi sociali con la trasformazione digitale, sia per quanto riguarda la trasformazione qualitativa (in negativo) del tempo specialmente per anziani, disoccupati ecc. Riconoscendo con Mückenberger che il nuovo *contratto sociale* deve «impedire efficacemente che le persone siano

⁶¹ R. SPAEMANN, *Persone. Sulla differenza tra "qualcosa" e "qualcuno"*, a cura di L. Allodi, Laterza, Roma-Bari 2005, 160.

⁶² R. MÜHLHOFF, *Die Macht der Daten. Warum künstliche Intelligenz eine Frage der Ethik ist*, Universitätsverlag Osnabrück-V&R Unipress, Osnabrück 2023, 58.

⁶³ Infatti anche Habermas parla di un'«insufficiente regolamentazione politica dei nuovi media», HABERMAS, *Nuovo mutamento*, cit., 27.

illegittimamente espropriate del loro tempo da parte di attori privati o pubblici o che abbiano un controllo diseguale (soprattutto gerarchico di genere) sull'uso del loro tempo»⁶⁴, bisogna ripensare le dinamiche sociali in una società in cui i modi di produzione e della gestione sociale del tempo libero saranno radicalmente cambiati.

Per quanto riguarda la concretizzazione di esigenze di protezione giuridica da realizzare nell'*infosfera*, si dibatte non solo intorno alla *privacy*, cioè all'integrità e alla riservatezza dei sistemi informatici, ma anche intorno ai diritti di accesso a Internet, all'oblio e all'anonimato su Internet, alla protezione dalla sorveglianza, alla proprietà digitale, alla spensieratezza digitale e all'educazione digitale, al diritto di tenere assemblee digitali e alla partecipazione politica e sociale, e, non ultimo, a un possibile diritto alla resistenza digitale. Ma quand'anche tali diritti fondamentali fossero codificati, si pone la domanda di come possano essere garantiti e sanciti contro il potere non più politico-statale (come vuole la dottrina classica dei diritti fondamentali) ma di imprese non territorialmente collocabili.

La Corte suprema europea tratta già aziende quali Google come se fossero legate dai diritti fondamentali (cf. il caso *Google vs Spain*)⁶⁵.

La necessaria riaffermazione del politico attraverso un *nuovo contratto sociale* deve, pertanto, essere in grado di riaffermarlo nei confronti di tale forte potere economico: «[i]l capitalismo della sorveglianza dipende dalla società, ed è solo attraverso un'azione sociale collettiva che la promessa di un capitalismo dell'informazione in linea con una terza modernità capace di prosperare può essere mantenuta»⁶⁶. Occorre contrastare la caratteristica di tale capitalismo della sorveglianza che organizza la società attraverso il «non-contratto» che «non è uno spazio per

⁶⁴ U. MÜCKENBERGER, *Metronome des Alltags. Betriebliche Zeitpolitiken, lokale Effekte, soziale Regulierung*, Sigma, Berlin 2004, 286.

⁶⁵ Cf. S. LEUSCHNER, *Eine "Charta der Grundrechte für die digitale Zeit", und warum wir sie brauchen*, in *VerfBlog*, 11 dicembre 2015, <https://verfassungsblog.de/eine-charta-der-grundrechte-fuer-die-digitale-zeit-und-warum-wir-sie-brauchen/> (accesso 1.12.2023).

⁶⁶ ZUBOFF, *Il capitalismo*, cit., 208.

relazioni contrattuali ma un'esecuzione unilaterale che rende superflue tali relazioni». Esso «desocializza il contratto, fabbricando la certezza e sostituendo con dei processi automatizzati promesse, dialogo, condivisione del significato, risoluzione dei problemi, risoluzione dei conflitti e fiducia»⁶⁷.

Riguardo alle esigenze appena descritte, bisogna infine domandarsi come descrivere il soggetto del *nuovo contratto sociale*, e a tal riguardo non si può non riferirsi all'emergere della specifica *vulnerabilità cognitiva* del soggetto nell'*infosfera*: essa consiste nel continuo oscillare «tra una mancanza di fiducia in se stesso e una fiducia eccessiva nell'informazione che riceve»⁶⁸. Anche per questa vulnerabilità i presupposti della stessa *società aperta* risultano messi in crisi dal successo dei *nuovi media*. Essa si traduce, però, nel paradosso etico che l'«essere monitorati è spesso accettato o addirittura praticato attivamente (*life-logging*). Le difese etiche contro la sorveglianza devono rispettare tali decisioni di auto-denuncia se non vogliono correre il rischio di essere respinte come moralismo»⁶⁹. Inoltre, il *paradosso della privacy* consiste nel fatto che le persone sono allo stesso momento consapevoli e addirittura preoccupate del rilevamento dei dati a larga misura, mentre non si adottano comportamenti di opposizione e non si è nemmeno disposti a "pagare" per rafforzarla⁷⁰. Una maggiore protezione della *privacy*

⁶⁷ *Ibid.*, 235.

⁶⁸ G. ORIGGI, *La democrazia può sopravvivere a Facebook? Egualitarismo epistemico, vulnerabilità cognitiva e nuove tecnologie*, in *Ragion pratica* 51 (2018) 2, 445-457, qui 448. «Avere ragione secondo i propri standard, rifiutare di farsi imporre qualsiasi standard epistemico percepito come una coercizione "dall'alto" è il tranello in cui cascano i cittadini più vulnerabili, che si sentono *empowered* cognitivamente e socialmente dal poter esprimere la loro opinione e non si rendono conto di essere vittime di nuove forme di controllo dell'opinione e di coercizione mentale», *ibid.*, 450.

⁶⁹ D. KAMMERER, *Überwachung.*, in J. HEESSEN (hrsg.), *Handbuch Medien- und Informationsethik*, Metzler, Stuttgart 2016, 191.

⁷⁰ «*Paradosso della privacy*. La stessa persona che afferma di essere preoccupata per la *privacy* non è disposta a spendere un soldo per essa. Piuttosto, rivela le informazioni private ai social media o ad altre piattaforme senza preoccuparsene troppo», GIGERENZER, *Perché l'intelligenza umana*, cit., 207.

dovrebbe mirare a una garanzia dell'autodeterminazione e dell'autonomia degli individui e non a forme di moralismo o paternalismo che spesso vengono suscitate dalla vulnerabilità cognitiva del soggetto o dai suoi comportamenti "paradossali". Franke analizza come questa tendenza sia diametralmente opposta al "principio libertà"; e occorre proteggere tale libertà per non cedere alle possibilità architettoniche delle nuove tecnologie quali si esprimono nel *social credit system*: per promuovere "sincerità" e "armonia" nella cultura, i punteggi di credito sociale sono stati introdotti, dopo la Cina, anche da altri Stati come la Thailandia, Myanmar, Vietnam, Venezuela e Tanzania⁷¹. Ma è anche chiaro che «[s]enza una regolamentazione da parte dello Stato di diritto, il principio della performance e la massimizzazione del profitto costringono a una sempre maggiore alienazione delle persone dai loro limiti costitutivi entro i quali possono plasmare le azioni del loro mondo di vita»⁷².

4. Conclusione

Per fondare il *nuovo contratto sociale* sulla garanzia delle basi sociali per la formazione di beni personali e relazionali indispensabili, non si può che partire dal fatto che «solo il riconoscimento fonda la

⁷¹ Cf. *ibid.*, 202. «Quelli con punteggi alti vengono premiati con benefici, come l'accesso a condizioni di credito più convenienti e controlli sanitari gratuiti, mentre a migliaia di persone con punteggi bassi viene negato l'accesso ad aerei e treni ad alta velocità, e ai loro figli viene precluso l'accesso alle scuole migliori. Le persone hanno cominciato a rimuovere "amici" dai social sulla base dei loro punteggi. [...] in Cina, il 94 per cento disse di aver modificato il proprio comportamento per influenzare positivamente il punteggio di credito sociale» (*ibid.*). L'autore constata però anche che in Cina la maggioranza è favorevole a tale sistema, «soprattutto quelli con un'istruzione superiore. Molti credono che fornisca una vera alternativa alla democrazia alimentando sincerità, armonia e crescita economica per tutti», *ibid.*, 203.

⁷² G. FRANKE, *Künstliche Intelligenz, Transhumanismus und menschliche Personalität*, Wissenschaftliche Buchgesellschaft, Darmstadt 2022, 339.

personalità»⁷³: da questo presupposto derivano concrete richieste al nuovo contratto sociale della società nella *digital matrix*. Le macchine non realizzano personalisticamente gli atti che identifichiamo come *actus humani* in quanto non li realizzano nella "maniera dello spirito", che afferma la realtà soltanto tramite il momento della negazione che costituisce l'oggetto del sapere come *opposto* a ciò che in questo momento emerge come *Io*⁷⁴. Nella misura in cui il "pensiero" viene concepito senza questa sua relazione intenzionale col «mondo fisico oggettivo», perché interpretato soltanto come «epifenomeni insignificanti», il mondo oggettivo-reale scompare⁷⁵. Poiché ai sistemi di intelligenza artificiale manca questa auto-referenzialità, in quanto privi di auto-coscienza e di auto-trasparenza, non si può in nessun modo attribuire ad essi "intelligenza"⁷⁶. Nella misura in cui essi realizzano "controllo tramite la trasparenza" in riferimento agli esseri umani, in se medesimi essi sono caratterizzati da opacità: nessun sistema di IA può dare motivi e ragioni per la giustificazione delle sue azioni.

Di fronte a questa sfida, Benanti solleva il problema del «bisogno di dire nuovamente l'umano: abbiamo bisogno di riuscire ad esprimere in parole adeguate alla comprensione dei nostri contemporanei il valore e l'unicità del nostro essere persona»⁷⁷, ma la questione è in che cosa ciò debba consistere. La risposta non può che rintracciarsi in un contrasto ai principi del transumanesimo ripartendo dalla natura dell'essere umano

⁷³ SPAEMANN, *Personne*, cit., 230.

⁷⁴ «Le persone non sono semplicemente ciò che sono. Esse sono determinate da una differenza rispetto a ciò che sono, attraverso un momento di negatività. La negatività distingue l'essere vivente dal non vivente. Le persone non provano soltanto sensazioni, non solo pensano, ma pensano anche un aldilà del pensiero [...]. Quando, nondimeno, diciamo che "il computer non pensa", intendiamo che non sa che cosa pensa. "Non sa" in realtà significa: non vive. Non vi è spirito nella macchina. Sapere è un modo di vivere», *ibid.*, 41-42.

⁷⁵ *Ibid.*, 53.

⁷⁶ Cf. S. KNELL, *Künstliche Intelligenz und menschliche Würde – ein aporetisches Verhältnis?*, in *Zeitschrift für Ethik und Moralphilosophie* 5 (2022) 2, 203-229, qui 219.

⁷⁷ P. BENANTI, *Digital age. Teoria del cambio d'epoca. Persona, famiglia e società*, San Paolo, Cinisello Balsamo (MI) 2020, 123.

che, attraverso il suo limite e la sua fragilità, realizza la sua razionalità e libertà, e ciò necessita di istituzioni di garanzia della sua libertà che sono i principi della libera concorrenza, da un lato, e della rappresentanza democratica, dall'altro. Socialmente e politicamente, la società deve essere integrativa e partecipativa, realizzando proprio in queste dinamiche quel riconoscimento costitutivo per l'essere umano (*Io-Noi*) che non può mai essere tecnicamente mediato.

MagIA!
L'INTELLIGENZA ARTIFICIALE
E LE NUOVE FORME DELLO SPIRITO

Luigi Vero Tarca

Abstract

L'IA, pur possedendo effetti "miracolosi", anzi proprio per questo, si trasforma in un gigantesco incantesimo dotato di effetti illusionistici per quanto riguarda il destino dell'umanità. Così si può parlare oggi di una magica trasformazione dell'energia psichica degli umani in una nuova forma di energia della quale noi non ci rendiamo nemmeno conto. Il problema centrale è allora quello della comprensione delle nuove forme dello spirito che progettano e governano tali immani trasformazioni energetiche, e quindi il possibile senso che tutto ciò verrà ad assumere.

Parole chiave: IA, Illusionismo, Religione tecno-scientifica, Nuove energie, Nuove forme dello spirito.

The AI, despite possessing truly "miraculous" effects, and indeed precisely for this reason, is transformed into a gigantic spell endowed with illusionistic effects regarding the destiny of humanity. Thus we can speak today of a magical transformation of the psychic energy of humans into a new form of energy of which we are not even aware. The central problem is then that of understanding the new forms of the spirit that plan and govern these immense energetic transformations, and hence the possible sense that all this will take on.

Keywords: AI, Illusionism, Techno-scientific religion, New energies, New forms of the spirit.

1. L'Intelligenza Artificiale (IA) come magia

1.1. *Magia come illusionismo e come miracolo*

Il titolo del mio intervento, mettendo in maiuscolo le due lettere finali della parola "MagIA", suggerisce una identificazione tra l'Intelligenza Artificiale (IA) e la magia. Questa parola ha un duplice significato, perché indica sia degli effetti illusionistici che ci fanno prendere per reale qualcosa che invece non lo è, sia dei fenomeni reali ma sorprendenti in quanto superano le nostre capacità e quindi sembrano realizzati da soggetti superiori agli esseri umani. Nel primo caso si tratta di *illusionismo*, nel secondo caso potremmo parlare di *miracoli*. Esempio paradigmatico del primo caso è lo spettacolo in cui il prestigiatore, dopo aver tagliato a metà la cassa in cui ha fatto entrare una donna, separa le due metà all'estremità delle quali compaiono rispettivamente il volto di quella donna e dei piedi femminili del tutto simili ai suoi: tutti *hanno visto* che quella donna è stata tagliata in due; dopo di che ella riappare "magicamente" integra. Il secondo caso è invece rappresentato, ad es., dalle situazioni in cui, grazie al telefono o a Internet, ci parliamo e addirittura ci vediamo come se fossimo nello stesso locale anche se in realtà ci troviamo a molti chilometri di distanza. Così la televisione e il nostro smartphone sono delle "sfere di cristallo" che però, come buona parte degli strumenti che usiamo oggi, compiono autentici *miracoli*, appunto perché producono eventi addirittura fino a prima impensabili.

È dunque importante distinguere questi due aspetti, quello fittizio e quello reale/innovativo, perché – anticipando un tema della conclusione – possiamo dire che la tecnologia, di cui l'IA è una sorta di punta di diamante, presenta entrambi gli aspetti: sfruttando gli effetti strabilianti delle prestazioni "miracolose" che offre effettivamente agli umani essa genera un'interpretazione complessiva della realtà che è largamente illusoria, prestandosi così a inganni di ogni tipo. Un caso paradigmatico è quello dei miracoli della tecnica in campo medico, grazie a interventi chirurgici sofisticatissimi (pensiamo anche solo alla laparoscopia) e comunque impensabili prima dell'utilizzo delle formidabili tecniche di

anestesia, ma poi anche grazie agli interventi diagnostici (radiografie) o farmacologici capaci di azioni cliniche precedentemente impensabili. E il discorso potrebbe estendersi a molte altre innovazioni. Ma proprio questa indiscutibile e meravigliosa potenza offre l'occasione di effettuare "operazioni" che vengono a godere del prestigio e della indiscutibilità propri della scienza pur avendo in realtà conseguenze (dirette o mediate) del tutto controverse e problematiche. È in questo senso che la tecnologia sfrutta un risultato reale per produrre un gioco illusionistico; con la differenza che, mentre quando va al circo l'umano è consapevole di avere a che fare con dei trucchi, quando invece si rapporta alla tecnologia, ad es. entrando nel mondo mediatico (televisione, Internet etc.), egli è convinto di trovarsi di fronte a qualcosa di reale, e quindi si comporta poi "realmente" in base ai giochi illusionistici di cui è preda. Al punto che già nel 1947 un pensatore come Wittgenstein non si sentiva di escludere che la tecnoscienza potesse rappresentare una catastrofe per l'umanità¹.

1.2. Il linguaggio come magia originaria

Si badi che questo non è un problema solo di oggi perché appartiene già alle origini della tecnica, basti pensare a quella tecnica originaria dell'umano che è il linguaggio. Voglio dire che (come Nietzsche aveva in qualche modo intuito) il linguaggio, che ha svolto un ruolo decisivo nel far compiere agli umani il salto evolutivo essenziale rispetto agli altri animali, è esso stesso intrinsecamente paragonabile a una sorta di gioco di prestigio. Nel senso che il suo tratto essenziale consiste proprio nel fatto che una parola costituisce un contenitore che può raccogliere al proprio interno una quantità pressoché infinita di oggetti con caratteristiche anche molto diverse se non addirittura opposte, circostanza che consente a chi lo gestisce di estrarre a suo piacere l'oggetto che di volta in volta preferisce. Da questo punto di vista il linguaggio può essere senz'altro paragonato al cappello a cilindro dal quale il prestigiatore

¹ L. WITTEGENSTEIN, *Pensieri diversi*, Adelphi, Milano 1980, 106.

estrae di volta in volta un libro, un coniglio, una bottiglia etc. In questi casi noi ridiamo e ci divertiamo, ma la faccenda diventa un po' più seria quando si tratta di ciò che di fatto guida la nostra mente e quindi la nostra vita.

Immaginiamo, ad es., un cappello a cilindro, sul quale è stampigliata, a forma di croce, la parola "cristiano". Se il "prestigiatore" ti chiede se sei cristiano, tu, pensando alle altissime parole di Gesù («Ama il prossimo tuo come te stesso, ama anche chi ti fa del male, cioè persino i tuoi nemici», e così via) rispondi convintamente di sì. A questo punto il prestigiatore estrae dal cappello un crociato, e ti arruola seduta stante nell'esercito in partenza per una crociata in Terra Santa; oppure estrae la Santa Inquisizione, e ti invita perentoriamente a seguirlo in Campo de' Fiori, a Roma, per prendere parte al rogo dell'eretico. Sia chiaro, sono esempi ultrasemplificati e quindi da molti punti di vista anche sbagliati, perché offrono un quadro fuorviante (in quanto tutt'altro che realmente "comprensivo") di situazioni estremamente complesse e difficili, ma che hanno il pregio di metterci immediatamente di fronte al paradosso per il quale noi umani ci troviamo spesso (quasi sempre...?) a comportarci in base a ragionamenti di questo genere: «lo amo i miei nemici, quindi li brucio vivi». Ragionamenti che noi stessi bocceremmo inappellabilmente se ci fossero proposti da uno studente anche ai primi anni di studio, e che tuttavia vengono tranquillamente accettati come guida della loro vita anche da individui che dichiarano una totale fedeltà alla ragione (comprese la logica e la scienza).

Si badi che questi "difetti" sono ben difficili da evitare perché essi dipendono direttamente da un carattere del linguaggio che ne costituisce proprio il pregio (valore/merito) essenziale e definitorio. Mi riferisco al fatto che la stessa parola può riferirsi a realtà anche molto diverse, cosa che consente appunto agli umani di applicare schemi/modelli linguistici fissi a situazioni della realtà molto varianti e permette loro, tra l'altro, di creare strutture sociali e forme organizzative estremamente flessibili, ampie e quindi potenti. Insomma, l'aspetto magico è già presente nel linguaggio stesso, e nella sua duplice veste: di fattore

realmente miracoloso, ma insieme anche di gioco illusionistico che può favorire ogni forma di menzogna e di inganno, quindi anche di prepotenza e così via.

2. La divinizzazione dell'IA e l'autodeposizione della tecno-scienza

Da questo punto di vista potremmo dire che il problema dell'IA non è l'artificiale, ma l'intelligenza; o, meglio, che se il problema è l'artificiale, allora proprio per questo lo è innanzitutto l'intelligenza. Così l'IA, che rappresenta un'accelerazione straordinaria del progresso tecnico, costituisce anche un rischio estremo per gli umani, come sta diventando sempre più chiaro a buona parte delle persone. In relazione ai problemi posti dall'IA gli umani vengono ad assumere degli atteggiamenti molto diversificati e spesso anche divergenti.

2.1. La divinizzazione dell'IA in quanto culmine del sapere scientifico

Un primo atteggiamento è quello della divinizzazione. Esso attribuisce all'IA un valore *etico* assoluto in quanto la considera il coronamento del sapere scientifico, inteso come la realizzazione compiuta dell'intelligenza/conoscenza e a sua volta vista come il pregio distintivo dell'essere umano, e in quanto tale l'IA viene di fatto vissuta come un'espressione del divino che rigenera e salva l'umano. Il motivo per il quale si attribuisce un valore etico indiscutibile all'agire tecnologico (cioè tecno-scientifico) è che esso, appunto in quanto scientifico, si basa su ciò che *vale universalmente*. In quanto tale, infatti, esso ha *valore* per tutti (in quanto resta da tutti affermato) e risulta quindi *innegabile* (dal momento che, essendo universale, resterebbe confermato persino da chi intendesse negarlo). In tal modo il sapere scientifico, e l'agire che ad esso si conforma, costituisce un *positivo* (dal momento che è qualcosa su cui tutti sono d'accordo) il quale, pur essendo limitato e parziale, è *assoluto*, dal momento che, essendo in-negabile (non-negabile), non può trasformarsi in un *negativo*. Così l'agire tecno-scientifico garantisce

l'accordo tra ogni essere umano e la natura nel suo complesso, compresi in particolare gli altri umani. In questo quadro la realizzazione di qualsiasi beneficio determinato-specifico per gli umani risulta essere un positivo perfetto in quanto libero/salvo rispetto al negativo.

È importante prestare attenzione al fatto che in tal modo l'agire tecno-scientifico possiede un valore etico (cioè è un bene per tutti) anche se non riguarda direttamente le questioni cui di solito ci riferiamo con questa parola ("etico"), cioè quelle personali/esistenziali, caratterizzate in senso psicologico, morale, spirituale etc. Ciò perché il suo valore essenziale consiste nella libertà/salvezza rispetto al negativo, garantita appunto dal suo essere in-negabile. Così, in quanto in-negabile, tale agire riguarda ciò che è *necessario* (*ne-cedo*: non cedo alle offese procurate dalla negazione e quindi non ne patisco i danni). E necessaria è la realtà in quanto *si ripete* costantemente, e per questo può prescindere da tutto ciò che è contingente, quindi soggettivo e perciò qualitativo etc. In tal modo il sapere scientifico di per sé prescinde da tutto ciò che riguarda scelte/preferenze relative al bene e al male, al piacevole e al giusto, quindi prescinde per principio da tutto ciò che è valoriale in senso soggettivo-particolare-qualitativo. Anzi, l'agire tecno-logico ha il pregio di valere per tutti *proprio perché* si limita a questo ambito oggettivo/necessario. Da questo punto di vista l'agire soggettivo ha valore solo nella misura in cui è conforme all'operare tecno-scientifico, cioè non entra in contrasto con questo. Proprio in quanto si limita a ciò che è necessario, la validità dell'atteggiamento scientifico è data *a priori*; così che lo scienziato non è tenuto, ogni volta, a chiedere il consenso di ogni individuo prima di agire. Come accade emblematicamente nel calcolo (la procedura "chiusa"): noi siamo sicuri che il risultato di 12.345×876 è $10.814.220$ *per tutti*, anche per coloro che non hanno mai fatto questo calcolo.

2.2. La parabola della tecno-scienza: dalla divinizzazione alla idolatria

Si avvia così il percorso di un sapere che – essendo considerato, in quanto innegabile, indiscutibile e infallibile – avvia una progressiva conquista della realtà in tutti i suoi vari aspetti, realizzando quel *progresso*

del quale l'IA costituisce il culmine, perché significa l'appropriazione, da parte della tecno-scienza, dell'elemento (l'intelligenza) che costituisce la premessa di ogni altro valore umano e quindi anche di ogni altra sua azione buona e giusta. È proprio la legittimazione/giustificazione "etica" di cui gode l'agire tecno-scientifico che rende non solo possibile, ma addirittura doveroso, per l'umano, estendere tale agire alla *totalità* della sua esperienza e quindi del mondo. Di qui, appunto, il valore assoluto/indiscutibile dell'IA, e la sua conseguente "divinizzazione". Ma il punto è che questa parabola – e proprio anche grazie al suo successo e alla sua trionfale estensione ormai totale – finisce per mettere in questione i suoi stessi fondamenti e quindi il suo stesso valore. Perché essa giunge al punto di mettere in crisi le stesse nozioni che la fondano e la legittimano, in particolare quella dell'"in-negabile" e quindi del "necessario".

Ciò accade lungo due versanti principali, quelli sui quali si è tradizionalmente fondato il valore universale del sapere scientifico: da un lato, ed essenzialmente, il *logos* (la *ragione* di cui tutti gli uomini sono dotati, come dice Aristotele: *anthrōpos zōon logon echōn*) dal quale scaturiscono gli infallibili procedimenti/calcoli logico-matematici; dall'altro lato i *sensi*, che, almeno nella misura in cui sono sostanzialmente uguali in tutti gli umani, garantiscono l'unanimità (quindi l'accordo) anche per quanto riguarda la conoscenza del mondo sensibile/naturale. Perché il sapere scientifico riguarda o fenomeni che tutti possiamo osservare senza modificarli (il movimento degli astri nel cielo) oppure esperimenti che noi stessi produciamo ma che comunque restano da tutti controllabili perché sono sempre ripetibili. Come del resto insegna il metodo galileiano, basato appunto su "dimostrazioni necessarie" e "sensate esperienze".

Ma tutto questo alla fine entra in crisi. Sul versante logico-matematico ci si rende conto che l'estensione totale del sapere scientifico non è possibile, perché allora esso dovrebbe riferirsi riflessivamente anche a se stesso, ma in questo caso diverrebbe contraddittorio: si veda in particolare Gödel (teoremi di incompletezza) – sulla scia di Russell (la classe di tutte le classi) e di Wittgenstein (i limiti della logica e l'esito mistico)

– ma poi anche Tarski (il paradosso del Mentitore), etc. Proprio perché tale metodo non può includere (e quindi fondare) se stesso, il limite di cui parliamo è *qualitativo* e quindi sostanziale, nel senso che alla fine *qualsiasi proposizione* resta infondata quanto al proprio valore innegabile. Io compendio tale circostanza in questa formula: «Il principio di non contraddizione è contraddittorio»; nel senso che, *se assunto come criterio universale del sapere*, tale principio risulta contraddittorio.

Ma anche sul versante empirico-esperienziale la legittimazione scientifica viene messa in questione. Perché i sensi dipendono sostanzialmente dalla natura umana, quindi dalla sua dimensione fisico-corporea, la quale a sua volta è uno degli elementi che può essere travolto dall'inesorabile processo del cambiamento fisico-naturale e quindi poi dalla stessa evoluzione zoologica. Del resto è proprio l'inarrestabile progresso tecno-scientifico che sta mettendo fuori gioco la "necessità" della natura umana; a cominciare dal fatto che oggi la conoscenza scientifica è sostanzialmente basata su artefatti tecnologici (osservatori astronomici etc.) piuttosto che su dati sensibili "immediati"; cosa che pone il problema della *fiducia* nell'operare del mondo scientifico. Insomma, oggi il ragionamento logico-matematico non è più innegabile, e la scienza fisica stessa – in particolare dopo il principio di indeterminazione, l'*entanglement* e così via – finisce per perdere la pretesa di costituire una *evidenza* che, ancorché parziale e provvisoria, è comunque universale e decidibile, quindi innegabile e oggettiva.

Tutto questo mostra come la fondazione ultima della legittimazione tecno-scientifica (cioè la convinzione "galileiana" che la natura sia scritta "in caratteri matematici", così che l'uomo è uguale al Dio creatore, sia pure solo intensivamente e non anche estensivamente) risulti in fondo una *fede*, piuttosto che una verità, e quindi come l'apparato e le istituzioni su di essa edificati costituiscano sostanzialmente una *religione*. Religione che, se non è consapevole di essere tale, si trasforma inesorabilmente in una *superstizione*, precisamente quella che possiamo chiamare la "superstizione tecno-logica" (ovvero tecno-scientifica). Circostanza, questa, confermata dal fatto che quelli della "scienza" ormai sono "misteri" dei

quali possono parlare solo i *sacerdoti* della tecnoscienza. Ed è appunto questa "religione" che legittima l'inarrestabile operare tecnologico e il suo carattere "totalizzante"; giacché la sostituzione dell'intera realtà naturale (compreso l'uomo) con entità di carattere artificiale è giustificata dal fatto che l'operare tecno-scientifico rispetta appunto la "lingua di Dio" (la matematica). Sicché ormai la tecno-scienza *garantisce* la *verità* (l'accordo tra il soggetto umano e il mondo) non tanto sostituendo le interpretazioni (ipotesi) falsificate con altre ipotesi corroborate, quanto piuttosto sostituendo la realtà non conforme ai nostri modelli (cioè "disubbidiente") con artifici costruiti in modo da garantire l'ubbidienza agli "ordini" della matematica. La scienza moderna (come già Kant aveva chiarito) riconosce come davvero reale solo ciò che (galileianamente) corrisponde alle qualità primarie, cioè all'aspetto quantitativo, e su questa base la tecnologia procede poi, implacabilmente, a sostituire *ogni* realtà con realtà "artificiali" (cioè funzionanti in maniera quantitativa). In breve, la scienza *seleziona* gli aspetti della natura considerando davvero reali solo quelli che corrispondono ai nostri modelli matematici, mentre la tecnica *sostituisce* la realtà con artefatti costruiti dall'uomo in base a siffatti modelli matematici.

È in questo modo che si intende superare anche la peculiare difficoltà legata al problema, da sempre noto, che la conoscenza scientifica presenta comunque un carattere essenzialmente aperto e quindi problematico, in quanto esposto alla possibile "smentita" da parte della natura (falsificazione). Perché tale problema, che nell'epoca pre-tecnologica viene risolto grazie al fatto che il valore "etico" della scienza è garantito dall'universale controllabilità degli esperimenti, nell'epoca tecnologica viene risolto sostituendo la natura con oggetti artificiali.

3. La verità dell'IA

3.1. L'IA come forma di "potere"

Venendo ormai meno, nell'epoca della tecnica "scatenata" (transumana, postumana) la stessa necessità (inviolabilità) della natura umana,

accade che persino il criterio "etico" della controllabilità delle teorie da parte degli umani venga risolto sostituendo gli umani con soggetti che si adeguano a ciò che le istituzioni tecno-scientifiche fanno loro credere. Così l'intera realtà naturale (oggettiva e soggettiva) viene rimpiazzata da una realtà artificiale: realtà "arbitrariamente" costruita dalla volontà/potenza umana, cioè dalla potenza della volontà umana, la sua *volontà di potenza*. Questo vuol dire che la religione di cui stiamo parlando è fondata sulla fede nella *potenza*, cioè nella capacità di far sì che il carattere *potenziale* dell'essere (*dynamis*), come tale aperto alla *possibilità*, si riveli *necessario* assumendo la forma dell'inesorabile *ripetizione secondo una legge*: l'eterno ritorno dell'identico (Nietzsche). In tal modo la vera natura della *necessità* (*anankē*) si mostra essere *dynamis*, ma nel senso di *possibilità* che si realizza determinandosi in forza della *potenza* propria della energia efficace (*energheia*). La quale poi, rispetto all'umano, si manifesta come fede nel *potere*, cioè nella capacità di operare in maniera in-contrastabile (in-vincibile) sull'umano stesso. In tal modo il sapere/potere diventa totalizzante e quindi, siccome alla fine si rivela essere infondato e quindi "ingiustificato", completamente totalitario.

La vera superstizione della nostra epoca è così quella tecno-cratice; che oggi è, di norma, tecno-demo-cratice. Prima della tecnica sovrumana (quella che si esercita direttamente sull'umano) il potere, ancorché "negativo", poteva essere accettabile e legittimo (quindi "positivo") in quanto i suoi svantaggi erano ampiamente compensati da grandi vantaggi (caso emblematico: il potere dell'esercito, che in parte opprime una popolazione, però la difende dalla violenza dei nemici). Naturalmente questo accade *a determinate condizioni*: che il potere resti all'interno di certi *limiti* (quale, innanzitutto, il rispetto della vita), cosa che appunto "definisce" il sistema democratico; ma poi a condizione che in ogni caso esso resti *originariamente* limitato dalla *intangibilità* della natura umana. A tali condizioni, anche se il potere commette errori, prepotenze etc., l'uomo resta sempre in grado di reagire e correggere queste "storture". Ma quando la tecno-crazia investe direttamente l'umano, la mancanza di giustificazione da parte del potere

mostra tutta la sua pericolosità. In particolare può accadere che potere tecno-scientifico e potere politico-democratico, pur essendo entrambi privi di un fondamento ultimo di valore e quindi in fondo "ingiustificabili", finiscano che dar vita a un gioco delle parti nel quale due soggetti entrambi infondati/"ingiustificati" pretendono di giustificarsi a vicenda: il politico giustifica le sue scelte scaricando sulla scienza la legittimazione delle stesse, e lo scienziato scarica sulle esigenze della "politica democratica" la giustificazione delle scelte fatte (politica che viene assunta, pensando a Churchill, come quella della quale non se ne dà una migliore, cioè meno dannosa).

La verità del sapere scientifico consiste dunque nella sua *capacità* (potere) di trasformare la realtà in qualcosa che corrisponde a ciò che *lo spirito* scientifico/intelligente vuole; consiste quindi in sostanza nella *potenza*, da parte dello spirito, di realizzare ciò che vuole.

Si badi che dire che la tecno-demo-crazia è una religione costituisce un'affermazione tutt'altro che offensiva, perché sottintende nei suoi confronti lo stesso atteggiamento di rispetto che viene riservato a tutte le religioni: da quella dei sacerdoti egizi (con le grandi opere che vengono ad essa attribuite: i lavori sul Nilo, le Piramidi etc.) a quella della Chiesa cattolica (pensiamo anche solo all'arte da essa generata), e così via. Pertanto dichiarare superstizioso l'atteggiamento fideistico di molti di coloro che nella religione attuale "credono" incondizionatamente e quindi ad essa si affidano è cosa lontanissima dal condannarla; come del resto riconoscere una superstizione – per esempio (in negativo) quella del gatto nero, oppure (in positivo) quelle del ferro di cavallo o dell'amuleto – è ben diverso dal prendersela con questi oggetti. In maniera un po' scherzosa, definire superstiziosa la pratica di chi, basandosi sull'osservazione che l'ombrello ci protegge, lo usa come paracadute gettandosi dalla cima di un palazzo (sfracellandosi) oppure lo usa per suonare il violino (straziando gli uditori), è cosa ben diversa dall'aizzare gli umani a distruggere gli ombrelli, cosa che tra l'altro farebbe perdere a loro (e ai loro violini) la protezione contro la pioggia. Analogamente, considerare superstiziosa la fede di chi ritiene che la tecno-scienza

garantisca il bene dell'umanità è cosa assolutamente lontana dallo scagliarsi contro i seguaci di questa fede-religione: semplicemente costituisce un invito a liberarsi dall'atteggiamento superstizioso che ormai la avvolge. Soprattutto perché in generale, come abbiamo visto, è una superstizione anche assumere l'intelligenza umana come qualcosa che è *incondizionatamente* un bene, essa infatti è un bene solo a *determinate condizioni*. Queste sono almeno due: da un lato che il soggetto umano del quale essa garantisce la soddisfazione (o della quale costituisce almeno un momento indispensabile) abbia davvero *potere* nei confronti della realtà; ma poi dall'altro lato, e in particolare, che tale trasformazione sia qualcosa di senz'altro *positivo*, cosa che presuppone pure l'assunzione che la natura umana sia definita da un'essenza immutabile e in sé buona.

Rendersi conto della "infondatezza" della scienza (Aldo Giorgio Gargani), e quindi della sua "ingiustificabilità" significa dunque rendersi conto (rievocando liberamente una felice espressione di Guglielmo Forni) che "il sogno è finito". Così la "divinizzazione" dell'IA (culmine della tecno-scienza) si mostra essere in realtà una forma di *idolatria*. Sicché l'apparato tecno-scientifico, proprio quando avanza la pretesa di presentarsi come manifestazione di natura divina, si trasforma nell'espressione di uno spirito maligno (luciferino/satanico). Si scopre che quella che abbiamo sempre chiamato "scienza" (intendendo con ciò un sapere eticamente valido perché controllabile da tutti) oggi non è più possibile e addirittura nemmeno immaginabile. Basti pensare anche solo alla circostanza che ormai tutto ciò che è rilevante per gli umani è sostanzialmente secretato (a livello innanzitutto militare ma poi anche industriale, quindi pure politico, mediatico etc.), sicché tutto ciò che la scienza afferma *presuppone/postula* la fiducia (*fede*) negli apparati che la producono e la governano. Così il suo valore "etico" si rovescia in dis-valore; per esempio, l'applicazione allo stesso fenomeno umano del metodo "per prova ed errore" (come accade inevitabilmente nei saperi che riguardano direttamente l'umano, quale è in particolare la medicina) porta alla fine dell'etica "umana" e quindi al rischio di un "film dell'orrore".

3.2. Il fine dell'IA è quello di sostituire l'umano piuttosto che di imitarlo

Quando ci si rende conto che la divinizzazione della tecnoscienza è una forma di idolatria, si rischia di cadere, per reazione, in una forma di demonizzazione che si apre a due atteggiamenti opposti: il primo è quello della "disperazione"; il secondo, per contro, è quella di una "fuga irrazionale": l'affidamento a una salvezza a priori di tipo soprannaturale (meta-fisico). La prima scaturisce dal rendersi conto dell'insuperabilità della trappola tecno-critica. Innanzitutto perché, dal punto di vista pratico, la tecno-scienza appare incontrastabile e quindi inevitabile/ineludibile. Per esempio, una volta che sono state inventate le armi da fuoco, nessuno, se vuole sopravvivere, può permettersi di non tenere conto di questa novità; operare questa rimozione sarebbe illusorio. Ma poi sarebbe perdente anche dal punto di vista "etico"; perché anche chi obietta all'IA che la sua giustificazione in fondo è solo il potere (il pre-potere), non è a sua volta in grado (fino a prova contraria) di esibire una posizione che non sia anch'essa una forma di contro-(pre-)potenza; sicché egli potrebbe avere davvero *ragione* solo se egli riuscisse ad avere *ragione* del mondo tecnico, cioè a sconfiggerlo; ma come può fare ciò rinunciando ad attrezzarsi con le armi fornite dalla tecnologia?

Per sfuggire a questa trappola e al suo esito disperante sorge spesso la tentazione di affidarsi a una via d'uscita consolatoria e quindi allettante, ma in realtà dannosa perché illusoria. Si tratta della convinzione che "possiamo stare tranquilli" perché comunque nemmeno il compimento totale della parabola tecno-scientifica (nel nostro caso, l'IA) è in grado di minacciare davvero il fenomeno umano, giacché *per principio* l'IA non può comunque giungere a *imitare* pienamente l'intelligenza umana. Il punto è che questa prospettiva, per quanto per molti aspetti vera, è ben lungi dal metterci al riparo dai rischi reali che comporta l'avvento dell'IA, perché un assunto centrale di tale prospettiva è che l'intelligenza umana non è comunque imitabile/riproducibile da parte di una macchina, giacché è radicata in un corpo vivente, grazie al quale si nutre di emozioni, sentimenti etc.; dal che segue che per evitare i rischi dell'IA è sufficiente portarla sotto il controllo degli umani. Ma questo

è illusorio, giacché l'uomo è il problema piuttosto che la soluzione: innanzitutto perché nemmeno la natura umana è un *dato immutabile*, e poi, soprattutto, perché nemmeno l'uomo si presenta come un *valore indiscutibile*.

Per comprendere tutto ciò bisogna incominciare a rendersi conto che – per dirla in una battuta – il fine reale dell'IA è, piuttosto che quello di *imitare* l'intelligenza umana, quello di *sostituire* tale intelligenza con altre forme di intelligenza, quelle che caratterizzeranno forme di vita *essenzialmente diverse* da quella umana. Perché il "fine" della tecnica è di sostituire *in toto* l'umano. Si capisce meglio questo punto se si osserva che i produttori di automobili certamente avevano come modello, o spunto, i carretti trainati da asini, muli e cavalli (i nativi americani chiamavano "cavallo di ferro" il treno), ma il loro fine non era quello di imitare tali animali alla perfezione (cioè in riferimento alla loro vita integrale: sensazioni, piaceri, dolori etc.), ma solo quello di riprodurne alcune caratteristiche. Così, il fatto che i cavalli non fossero "riducibili" a macchine non ha impedito agli umani di sostituirli con altri mezzi di trasporto, determinando in tal modo una sorta di estinzione di quei meravigliosi animali. Pertanto oggi, anche ammettendo che nessuna *macchina* potrà mai riprodurre integralmente gli umani, bisogna capire che ciò è ben lungi dal garantire che lo spirito che governa la tecnica non condurrà a una estinzione degli umani, considerati ormai antiquati e quindi superflui (come abbiamo incominciato a capire: da Günther Anders a Yuval Noah Harari).

Si dirà: ma comunque solo degli individui umani possono essere in grado di compiere un'operazione del genere; sicché, qualunque "miracolo" compia l'IA, esso non potrà cancellare *in toto* il fenomeno umano, perché lo presuppone. Come sempre, in un certo senso anche questo è vero; ma si tratta di capire in che cosa consista, esattamente, siffatta dimensione *intangibile* dell'umano, che potremmo chiamare "spirituale". Il punto è che il soggetto/spirito che, tramite la tecnica (sic!), sostituisce l'umano, può dare vita a forme di vita anche lontanissime da quelle che noi siamo soliti indicare e riconoscere come "*homo*". In

fondo anche il fenomeno umano può essere considerato un fenomeno animale, ma sarebbero state poco avvedute le scimmie o i polli o gli uccelli che avessero detto: in fondo solo una forma di vita animale può sbarazzarsi di noi, quindi non abbiamo nulla da temere dall'avvento dell'umano... Voglio dire che l'operare tecnico può condurre a sostituire il fenomeno umano con forme di vita completamente differenti da esso, anche se un'operazione siffatta richiederà che i nuovi soggetti conservino *qualche* tratto di quelli che oggi chiamiamo esseri umani, precisamente quei tratti che i padroni postumani della vita riterranno degni di essere conservati; dunque anche le emozioni e i sentimenti, eventualmente, ma comunque a loro insindacabile giudizio.

Sto dicendo che, essendo consolatorio/illusorio credere che il fine dell'IA sia semplicemente quello di *imitare* l'intelligenza umana, essa IA può venire usata al fine di costruire *ex nihilo* (o *ex Nilo*, per richiamare scherzosamente il carattere religioso dell'apparato tecno-scientifico) forme di vita, destinate a *rimpiazzare* gli umani e a dominare la terra, che faremmo fatica a chiamare ancora "umane" benchè scaturiscano dalla vicenda umana. Se l'umano riuscirà a riprodursi, ciò potrà accadere solo grazie a una benevola concessione dei padroni della vita sulla terra. Del resto, con che titolo (e sulla base di quale potere) gli umani "naturali" pretendono di avere il *diritto* riprodursi?

4. Le nuove forme dello spirito

4.1. La necessità della libertà

Il problema che si impone (*epi-histēmi: epistēmē*), a questo punto, è quello di comprendere quali siano le forme dello spirito che gestiscono la tecno-crazia; cosa che a sua volta è possibile solo da parte di una forma dello spirito che sia capace di alzare lo sguardo all'altezza di quello del soggetto che cerca di comprendere.

E questo richiede che lo spirito acceda a una nuova concezione della verità, quella che gli consente di passare dalla verità *in-negabile* a quella

onnaletica (onn-com-positiva)². Del resto, è la nozione stessa dell'in-negabile a implicare e quindi a evocare la dimensione del tutto-positivo assolutamente salvo rispetto al negativo. Tale dimensione si trova nel campo della *potenzialità* piuttosto che in quello della *necessità*, ma nel senso che è capace essa stessa di determinare il campo delle possibilità. Ed è proprio in questo regno della *dynamis* (potenza) che diventa possibile distinguere il *potere* (forza che è anche pre-potenza) dalla *efficacia*. Si schiude qui, in particolare, la dimensione nella quale ciò che è davvero buono risulta effettuale/effettivo proprio nella misura in cui è diverso da ciò che ha (pre-)potere (come accade, emblematicamente, nel caso dell'armonia/amore). Questo esige il passaggio dall'Intelligenza Artificiale all'Intelligenza Artistica, cioè "con-positiva"; anch'essa *artificiale* (in quanto prodotta dal vivente), ma generata seguendo il canone della bellezza piuttosto che quello della potenza, e quindi della *efficacia* piuttosto che del *potere*.

Questa dimensione, *altra* rispetto a quella della necessità, è tuttavia implicita nella stessa sfera della necessità. Perché questa in realtà si limita ad aprire una serie di *possibili* opzioni (le possibili mosse degli scacchi, oppure la possibilità di portare l'acqua a 100° per farvi bollire qualcosa) rispetto alle quali la realtà effettuale determinata dalla *volontà* è sempre qualcosa di *eccedente*. Perché la necessità dell'*implicazione* logico-matematica (pensiamo per esempio all'operazione "2 x 4 = 8") si costituisce come un *condizionale*: se fai *p* (2 x 4) allora ottieni *q* (8); ovvero anche: a *condizione* che sia dato *a* (2) e che tu voglia ottenere (8) devi fare *b* (moltiplicare 2 per 4). Ma la decisione di *fare* qualcosa (qui, moltiplicare per 4) piuttosto che qualcos'altro resta sempre una scelta che per definizione *eccede* il regno della necessità. Del resto, e più radicalmente, basti riflettere sul fatto che ogni *ripetizione* è sempre,

² Chi volesse approfondire la prospettiva filosofica alla quale faccio qui riferimento può incominciare a leggere, dei miei scritti, *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. Màdera), Mondadori, Milano 2003; *Verità e negazione. Variazioni di pensiero*, a cura di Th. Masini, Cafoscarini, Venezia 2016.

per un altro verso, qualcosa di nuovo, quindi una *innovazione*. Sicché (ispirandomi liberamente a Raimon Panikkar) potrei dire che *in verità* la legge della realtà (legge che in quanto tale esprime una *necessità*) è la *libertà*: la realtà è *necessariamente libera*.

In questa nuova dimensione veritativa emerge che ciò che per l'umano è davvero rilevante comprende essenzialmente tale parte che eccede la necessità/ripetizione. Lo stesso sapere in-negabile esige dunque un meta-sapere: quello del *senso* di ciò che accade, e quindi dello *spirito*, che schiude la dimensione di ciò che è valoriale e qualitativo. Solo in questa dimensione è possibile tentare una risposta alla domanda circa il valore (etico) del "progresso" tecnologico. Domanda, questa, davvero difficile, perché si deve tenere presente che proprio l'istanza etica fondamentale esige un qualche superamento dell'umano (il vivente mortale e mortifero), superamento che però rischia di peggiorare il male che cerca di combattere. Perché, da quale pulpito l'umano si permette di *giudicare* e *condannare* la tecnica? L'interrogativo che in tal modo si pone è quello relativo ai progetti delle forme di vita destinate a sostituire l'umano, e quindi alla natura dello spirito (del soggetto) che le crea, le genera e le vivifica. Quali sono i progetti circa queste nuove creature? E a quali forme spirituali essi corrispondono? Tutto quello che, sulla base di quanto ci è dato sapere, possiamo capire, è che anche l'uomo verrà sostituito da qualche artefatto capace di corrispondere ai "Suoi" progetti. Ma "Suoi" di chi?

4.2. La libertà rispetto alla necessità del negativo

Lascio qui aperta tale domanda, la quale indica la grande questione che a questo punto si spalanca, e mi limito a fornire uno spunto di riflessione. La verità che qui si annuncia può davvero superare l'in-negabile solo se è capace di superare ogni negazione (*omnis negatio est contradictio*), e può ottenere tale risultato solo realizzandosi come *puro positivo*, cioè mediante *la pura differenza*. Solo questa, infatti, può liberare la vita rispetto alla *necessità del negativo*, perché questa costituisce una trappola ineludibile finché si resta all'interno dell'equivalenza tra

differenza e negazione, dal momento che (a questa condizione) anche ciò che intende essere *diverso* dal negativo si presenta per ciò stesso come a sua volta negativo, precisamente nei confronti del negativo; perché "Il negativo del negativo è negativo".

In una battuta potrei dire: poiché in questa dimensione *tutto è possibile*, è possibile liberarsi anche dalla *necessità* che rende insuperabile il negativo, dal momento che il negativo è definito come ciò che implica la propria negazione, cioè una risposta contraria e *necativa* (da *nex*, *necis*: morte, uccisione). Solo un'ottica libera rispetto alla necessità del negativo può darsi una forma di vita capace di riproporre il valore ultimo dell'umano e della sua intelligenza (armonia cosmica) in una maniera libera rispetto alla necessità "mortale" del fenomeno umano. Perché il tentativo di eliminare il danno-dolore può condurre a riprodurre, estendere e intensificare il danno stesso. In una battuta: se il problema dell'umano è (pensando qui al buddhismo) il dolore/morte, allora l'unica soluzione radicale (logica/necessaria) di questo problema (all'interno dello sguardo *necativo* del mortale) è quella di eliminare tutti i mortali. Vuoi eliminare la morte? Semplice: elimina tutti i mortali. Vuoi garantire l'eliminazione della violenza umana? Semplice, elimina tutti gli umani. *All'interno di quest'ottica negativa* la prospettiva più ottimistica per l'umano è che il futuro assuma la forma del suicidio assistito delle forme di vita umane. Ma dal punto di vista puramente positivo è possibile anche un'altra soluzione (un'altra forma di vita), quella che pensa a una "trasfigurazione" delle forme di vita umane: il problema della morte e del dolore viene affrontato in maniera diversa da quella che consiste nel negare/*necare* ciò che causa morte e dolore. Ma in che modo questo è possibile? Quali soggetti spirituali possono compiere questo "miracolo"? In merito a ciò dobbiamo incominciare a dare un'occhiata attenta a tutta la tradizione sapienziale mondiale (orientale e occidentale), comprensiva delle religioni, delle culture, dell'arte etc.

E certo i problemi che qui si pongono sono enormi. A cominciare da quello che solleva la questione della *proprietà* degli strumenti tecnologici, cioè dell'apparato tecno-scientifico, che è ormai un immenso

campo di produzione di armi (l'uomo come produttore di ordigni: Italo Svevo). Questione peraltro connessa a quella relativa al carattere "pubblico" del sapere tecnologico, la quale pone il problema del superamento del *segreto* che definisce ormai l'intelligenza umana e in particolare quella artificiale, che scherzosamente potremmo qui chiamare l'"Intelligence Artificiale" per evidenziare cioè il fatto che ormai l'intera intelligenza umana è nelle mani dei Servizi segreti di origine tecno-militare. Una risposta soddisfacente a tali questioni richiede un pensiero assolutamente creativo e, in questo senso, pienamente artistico: un atto davvero *miracoloso*.

5. Il vero miracolo dell'IA: dall'incantesimo illusorio al meta-miracolo della ricreazione dello spirito

Con queste parole, che rievocano i miracoli, ritorniamo infine alla questione iniziale, quella della magia. La tecno-scienza, abbiamo detto, fa autentici miracoli, perché rende reali cose che senza di essa resterebbero pure fantasie, semplici "sogni". Con ciò essa consente di distinguere i miracoli (gli effetti incredibili ma reali) dalle fantasie (l'illusionismo, quindi i trucchi e anche gli inganni) e in questo senso costituisce *un autentico progresso*. Ma per altri versi, e tuttavia proprio per ciò, essa è anche un gigantesco gioco di prestigio illusionistico: fa credere che il suo (incontestabile e perciò indiscutibile) *progredire* garantisca in quanto tale il *miglioramento* della vita umana nel suo complesso. Invece, affinché questo accada davvero, è necessario che lo spirito che governa la tecno-scienza sia benevolo nei confronti degli umani, e ciò è tutt'altro che scontato. Questo è appunto il nostro compito di uomini del nostro tempo: generare e "ricreare" lo spirito che *disegna* il mondo del tutto-positivo all'interno del mondo della tecnica, il mondo del "dopo l'uomo".

La tecno-scienza, dunque, è un fenomeno che presenta entrambi gli aspetti della magia, sia quello "miracoloso" che quello "illusionistico".

Potremmo dire che la tecnica "scatenata" usa dei miracoli effettivi (Internet, i robot che imitano, superano e sostituiscono gli umani) per compiere un gioco di prestigio (una "magia", un incantesimo) che determina una trasformazione radicale della vita sulla terra, venendo così a compiere un gesto che può costituire un "miracolo" di secondo ordine (meta-miracolo), che però può costituire, per gli umani, anche un incubo, un autentico inferno; perché a entrare in gioco non sono solo i post-umani, ma anche gli extra-umani e quindi pure, almeno potenzialmente, i dis-umani (anti-umani e contro-umani). Insomma, l'IA è una magia che sicuramente produce miracoli, ma nello stesso tempo essa fa incantesimi; i quali rischiano di rovesciarsi in *inganni* catastrofici.

Dovendo chiudere, ricorro a un paragone per dare una prima immagine di quanto sto cercando di comunicare. Pensiamo al "miracolo" dell'energia elettrica: la tecnica umana converte l'energia idrico/idraulica in energia elettrica; le gocce d'acqua, "condotte" nelle centrali elettriche, con la loro "caduta" illuminano tutte le città del mondo. Magia! Ma magia reale, cioè miracolo. Ebbene, che cosa sta accadendo oggi? La nostra energia psichica, quella che caratterizza la vita umana (comprensiva della nostra affettività, cioè di quello che siamo *in verità*, nel profondo, non solo nel nostro essere calcolanti/calcolatori), viene *interamente* raccolta in condotte e portata in laboratori/centrali (il meta-mondo virtuale) che la trasforma in qualcosa che noi non riusciamo non solo a vedere ma nemmeno a immaginare. Ogni giorno la nostra energia psichica (qui l'equivalente dell'acqua) viene trasformata in altro. Ma in che cosa? In una nuova, definitiva arma (come, a suo tempo, è successo con la bomba atomica) o in qualcosa di completamente diverso? E non potrebbe anche essere qualcosa grazie a cui a farsi luce è proprio la nostra stessa esistenza? Certo anche questo è possibile, ma richiede un miracolo; anzi, propriamente un autentico super-miracolo (o meta-miracolo): quello che realmente trasforma l'illusionismo tecno-logico (cioè l'inganno-catastrofe) in una *reale* trasformazione "armonica" della vita umana. L'esito illusionistico-catastrofico è necessario (inevitabile) nella misura in cui si resta all'interno di una logica *polemica*. In effetti, quella

che oggi viene chiamata "scienza" è in buona parte un apparato tecnologico che è sostanzialmente rivolto alla costruzione di armi, e rispetto al quale gli effetti benefici sono semplicemente delle ricadute (*fall out*) di carattere sociale. In maniera schematica ma paradigmatica: Internet è la ricaduta sociale di quella sorta di Progetto Manhattan permanente in cui consiste orma la vita politico-istituzionale degli U.S.A.

Lo spirito che scorge queste cose sa vedere qual è il miracolo effettivo che sta accadendo con la (grazie alla) *nostra* energia psichica. Ma noi normalmente non lo vediamo, almeno finché restiamo solo gocce d'acqua che non comprendono questa dinamica; e soprattutto finché ci facciamo incantare dal gioco illusionistico che ci fa credere che l'importante di quello che sta accadendo sia il delirio narcisistico-egocentrico con il quale ogni individuo può sparare in giro tutto quello che gli salta in mente. Ma laddove vi è spirito, lo spirito fiuta, capisce e resta quindi in gioco rispetto a quello che sta accadendo davvero. Ed è quindi in grado di raccogliere la grande domanda che, a questo punto, chiede a ciascuno di noi: a quale forma dello spirito ti senti appartenere? A quale forma dello spirito tu appartieni?

INTELLIGENZA ARTIFICIALE PER LEGGE ORIGINI E SVILUPPI DEL TECNODIRITTO CONTEMPORANEO

Paolo Moro

Abstract

L'IA sta attuando trasformazioni economiche e sociali con un'accelerazione tale da indurre il diritto europeo ad approntarne una regolazione con un'apposita legge (*Artificial Intelligence Act*). Le principali nozioni (legale, tecnologica e ibrida) dell'IA, che si rappresenta come paradigma innovativo dell'odierno tecnodiritto, sono esaminate nella prospettiva teorica dell'informatica giuridica, rilevando la possibilità di attribuire una vera e propria "personalità elettronica" a robot sofisticati, tenendo conto della criticità del paragone tra esseri umani e macchine.

Parole chiave: Intelligenza artificiale, *Artificial Intelligence Act*, Tecnodiritto, Personalità elettronica, Pensiero ibrido.

AI is carrying out economic and social transformations with a so invasive acceleration as to induce European law to prepare a thoughtful regulation by the new Artificial Intelligence Act. The main notions (legal, technological and hybrid) of AI, which is represented as an innovative paradigm of today's technolaw, are illustrated in the perspective of Legal Informatics, noting the possibility of attributing a real "electronic personality" to sophisticated robots, taking into account the criticality of the comparison between humans and machines.

Keywords: Artificial Intelligence, Law and Technology, hybrid thinking, electronic personality

1. Definizione legale

L'ampia diffusione dell'intelligenza artificiale, alimentata dai rapidi progressi della tecnologia informatica, ha reso cruciale in Europa non solo la riflessione etica, ma anche l'urgente necessità di stabilire norme giuridiche per regolare questo fenomeno, al fine di garantire lo sviluppo continuo del settore industriale e di proteggere simultaneamente il valore intrinseco della persona in base a principi generali di tutela della dignità¹. Prendendo atto delle implicazioni umane ed etiche connesse alle principali tecnologie emergenti ed esponenziali nell'era digitale, l'Unione Europea ha introdotto un rilevante rinnovamento della legislazione in materia di intelligenza artificiale, dopo il fondamentale intervento normativo del Regolamento Generale sulla Protezione dei Dati n. 679 del 2016 (GDPR).

Le azioni di studio e di discussione giuridica e politica che sono state intraprese e che si concluderanno dopo un decennio hanno avuto e avranno lo scopo di istituire una legislazione in grado di affrontare in modo appropriato le sfide etiche e sociali poste dall'IA e di bilanciare il progresso tecnologico con la tutela dei diritti e dei valori fondamentali dei cittadini europei. In sostanza, l'Europa unita autorizza e nel contempo limita l'IA per legge. Già nella premessa della risoluzione del 16 febbraio 2017, recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, il Parlamento europeo ha constatato che «l'umanità si trova ora sulla soglia di un'era nella quale robot, bot, androidi e altre manifestazioni dell'intelligenza artificiale sembrano sul punto di avviare una nuova rivoluzione industriale, suscettibile di toccare tutti gli strati sociali, rendendo imprescindibile che la legislazione ne consideri le implicazioni e le conseguenze legali ed etiche, senza ostacolare l'innovazione».

¹ L. FLORIDI - J. COWLS - M. BELTRAMETTI - R. CHATILA - P. CHAZERAND - V. DIGNUM - C. LUETGE - R. MADELIN - U. PAGALLO - F. ROSSI - B. SCHAFER - P. VALCKE - E. VAYENA, *AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds & Machines*, 28/4 (2018) 689-707.

Dopo una serie articolata di lavori preparatori, il 21 aprile 2021 la Commissione Europea ha presentato la proposta di regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (*Artificial Intelligence Act*). Il regolamento, che è un atto normativo di portata generale e che sarà obbligatorio in tutti i suoi elementi e direttamente applicabile in ciascuno degli Stati membri (in base all'art. 298 del Trattato sul Funzionamento dell'Unione Europea), giungerà all'approvazione nel 2024 ed entrerà in vigore due anni dopo, diventando la prima legge generale del mondo sull'IA.

È probabile che l'esigenza regolatoria *ex lege* di un fenomeno tecnologico ormai diffuso obbedisca alla mentalità dei sistemi di *civil law*, come quello dell'Europa continentale, e che negli ordinamenti di *common law* (tra i quali gli Stati Uniti e il Regno Unito) prevalga la tutela giudiziaria dei diritti e degli interessi coinvolti. Ciò che importa rilevare in questa sede, però, è che il testo legislativo non contiene una definizione specifica di "intelligenza artificiale". L'art. 3 della proposta designa il «sistema di intelligenza artificiale» come un *software* che può, per una determinata serie di obiettivi definiti dall'uomo, generare effetti (*output*) quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagisce.

Nella motivazione del progetto si precisa che la definizione di «sistema di intelligenza artificiale», contenuta nell'art. 3 della proposta, mira ad essere il più possibile neutrale dal punto di vista tecnologico e adeguata alle esigenze future, tenendo conto dei rapidi sviluppi tecnologici e di mercato relativi al fenomeno. L'allegato I della proposta afferma che la definizione potrà essere aggiornata dalla Commissione in relazione all'evoluzione tecnologica del *software* che, integrando il sistema di intelligenza artificiale, è oggi sviluppato con tecniche di apprendimento automatico, approcci di rappresentazione simbolica della conoscenza e metodi statistici e di ottimizzazione.

Tenendo conto che nel *Digesto* giustiniano è riportato il noto avvertimento di Giavoleno, secondo cui «ogni definizione nel diritto è

pericolosa» poiché «basta poco per sovvertirla»², è interessante notare come la Commissione Europea abbia ritenuto opportuno evitare una definizione permanente di “intelligenza artificiale” e di considerarla di principio adattabile nel tempo, in ragione del perdurante cambiamento tecnologico.

Le nuove regole, che presuppongono una nozione evolutiva della norma che definisce un fenomeno in costante e progressivo cambiamento, saranno applicate direttamente e nello stesso modo in tutti gli Stati membri e seguono un approccio basato sul rischio: infatti, il nuovo regolamento pone alcuni divieti nonché alcuni limiti espliciti per le applicazioni di IA considerate ad alto rischio. Dunque, la proposta di regolamento considera vietate pratiche di manipolazione del comportamento attraverso tecniche subliminali o di sfruttamento delle vulnerabilità di minori o disabili o di profilazione illegittima; attribuzione di un punteggio sociale basato sull’AI con trattamento sfavorevole per le persone da parte di autorità pubbliche; ricorso a sistemi di identificazione biometrica remota “in tempo reale” in spazi accessibili al pubblico a fini di attività di contrasto, salvo talune eccezioni.

Inoltre, il nuovo AI Act stabilisce obblighi riguardanti i dati, la documentazione e la tracciabilità, la fornitura di informazioni e la trasparenza, la sorveglianza umana nonché la robustezza e la precisione per fornitori e utenti dei seguenti sistemi di AI considerati ad alto rischio: identificazione e categorizzazione biometrica delle persone fisiche; gestione e funzionamento delle infrastrutture critiche; istruzione e formazione professionale; occupazione, gestione dei lavoratori e accesso al lavoro autonomo; accesso a prestazioni e servizi pubblici e a servizi privati essenziali e fruizione degli stessi; attività di contrasto, gestione della migrazione, dell’asilo e del controllo delle frontiere; amministrazione della giustizia e processi democratici. Per tutti i fornitori di sistemi di AI a rischio minimo, invece, è stabilita la possibilità di seguire un codice di condotta istituito autonomamente con impegni volontari relativi, ad

² «*Omnis definitio in jure civili periculosa est: parum est enim ut non subverti possit*». Cf. GIAVOLENO in D. 50, 17, 202.

es., alla sostenibilità ambientale, all'accessibilità da parte delle persone con disabilità, alla partecipazione dei portatori di interessi alla progettazione e allo sviluppo dei sistemi di AI, nonché alla diversità dei gruppi che si occupano dello sviluppo.

2. Nozione tecnologica

Le inevitabili e sempre più consistenti interazioni tra esseri umani e sistemi di IA nelle relazioni sociali, già particolarmente evidenti nel mondo del lavoro, si stanno accrescendo sempre di più con lo sviluppo scientifico e tecnologico, soprattutto nei settori più avanzati, come la bionica e la robotica³. Queste relazioni appaiono non solo collaborative e vantaggiose per l'umanità o per l'ambiente, ma risultano anche conflittuali, con la conseguente esigenza di stabilire un nuovo codice etico e legale di un'intelligenza artificiale destinata ad aiutare l'umanità⁴, fissando anche metodi giuridici di risoluzione delle controversie generate non solo dal contrasto tra soggetti umani, ma tra questi e le più sofisticate macchine intelligenti.

Come è noto, il sintagma "intelligenza artificiale" è una metafora che, come gran parte delle dirompenti invenzioni che oggi provengono dall'informatica (per esempio: il *web* o il *cloud*), è una figura retorica che tenta di spiegare un fenomeno ignoto con un termine noto nell'esperienza⁵. Il termine nasce il 31 agosto 1955 negli Stati Uniti, durante gli studi della scienza dei calcolatori elettronici (*computer science*) e della cibernetica (*cybernetics*), attualmente confluiti in vario modo

³ P. MORO, *Bio robotics and fundamental rights. Issues and limits of artificial intelligence*, in D. PROVOLO - S. RIONDATO - F. YENISEY (eds), *Genetics, Robotics, Law, Punishment*, Padova University Press, Padova 2014, 517-531.

⁴ D. BUTLER, *AI summit aims to help world's poorest*, in *Nature*, 546, (6 /2017) n. 7657, 196-197.

⁵ C. SARRA, *Lo scudo di Dioniso. Contributo allo studio della metafora giuridica*, FrancoAngeli, Milano 2010.

nell'informatica e nella robotica, quando lo scienziato nordamericano John McCarthy propone ad alcuni colleghi non meno brillanti (come Marvin Minsky, Nathaniel Rochester e Claude Shannon) di designare con questa metafora il titolo di un progetto estivo di ricerca relativo a un modello computerizzato della mente umana, che sviluppava i presupposti della cibernetica del matematico Norbert Wiener⁶.

Nel *paper proposal* del 1955 sono già presenti i tratti essenziali dell'intelligenza artificiale contemporanea: automazione e simulazione delle funzioni del cervello umano; utilizzo del linguaggio naturale; creazione di reti neurali; teoria della dimensione del calcolo; auto-miglioramento delle macchine; astrazioni dai dati sensoriali; casualità e creatività. Il fenomeno non è dunque affatto nuovo, ma si è diffuso ampiamente nelle analisi multidisciplinari delle scienze morali e sociali, tra le quali l'etica e il diritto, anche a causa dell'evoluzione esponenziale delle più importanti tecnologie digitali dell'epoca contemporanea. Nella scienza informatica odierna, infatti, l'IA è considerata «una disciplina che studia i fondamenti teorici, le metodologie e le tecniche che permettono di progettare sistemi *hardware* e sistemi di programmi *software* capaci di fornire all'elaboratore elettronico delle prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana»⁷.

Peraltro, l'intelligenza artificiale è l'area dell'informatica in cui più si è sviluppata la riflessione filosofica, in ragione della evidente e radicale questione riguardante la natura dell'intelligenza umana e la conseguente possibilità di stabilire se un sistema artificiale (una macchina) sia dotato di tale qualità soggettiva⁸. Come afferma Pat Hayes nel suo

⁶ J. MCCARTHY - M. L. MINSKY - N. ROCHESTER, and C.E. SHANNON, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955*, in *AI Magazine*, vol. 27, n. 4 (12/2006), 12.

⁷ M. SOMALVICO, *L'Intelligenza Artificiale*, Rusconi, Milano 1987.

⁸ A. MONTANARI, *Questioni di tecnoetica in intelligenza artificiale, robotica e bionica*, in P. MORO (cur.), *Etica informatica diritto*, FrancoAngeli, Milano 2008, 33-50.

*Naive physics manifesto*⁹, ciò che si vuole rappresentare con le applicazioni dell'intelligenza artificiale è il modo in cui le persone vedono il mondo, non il mondo come viene visto dalla fisica, con la conseguente necessità di elaborare modelli che riguardano non solo oggetti o eventi, ma anche percezioni e azioni inerenti alle modalità di comprensione del mondo da parte delle persone. Data questa situazione in continua evoluzione, dunque, è possibile notare che la prudente definizione giuridica del fenomeno in esame, prospettata dal nuovo AI Act, è nata da una nozione tecnologica delle macchine intelligenti, che designano oggi sistemi meccatronici elaborati da programmi *software* agenti nella realtà virtuale o integrati da *hardware* che, come si nota nell'esperienza sociale, non sono necessariamente dotati di morfologia androide¹⁰. L'esigenza di una disciplina legale, infatti, deriva non tanto dalla logica degli algoritmi o dal ragionamento meccanico, ma dalle applicazioni dell'intelligenza artificiale che emulano comportamenti esteriori e che producono effetti intersoggettivi, giacché in epoca moderna il diritto si distingue dalla morale perché regola essenzialmente relazioni di rilevanza sociale.

In letteratura i robot sofisticati, capaci di eseguire calcoli complessi e di esibire condotte reattive in risposta a stimoli esterni, sono stati considerati soggetti morali e centri di autodeterminazione responsabile, con conseguenze problematiche soprattutto nel diritto penale¹¹. La possibilità di attribuire una personalità giuridica ai robot particolarmente evoluti e sempre più simili agli umani è già acquisita nel dibattito filosofico e scientifico, nel quale si applicano categorie che tradizionalmente identificano la tutela del soggetto umano (come i diritti soggettivi) ai

⁹ P.J. HAYES, *The naive physics manifesto*, in *Expert Systems in the Micro-Electronic Age*, Edinburgh University Press, Edinburgh 1978.

¹⁰ P. MORO, *Intelligenza artificiale e tecnodiritto. Fondamenti etici ed innovazione legislativa*, in P. MORO (cur.), *Etica, diritto e tecnologia. Percorsi dell'informatica giuridica contemporanea*, FrancoAngeli, Milano 2021, 7-24.

¹¹ 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.

sistemi di intelligenza artificiale¹². Ne prende atto anche la risoluzione del Parlamento Europeo del 16 febbraio 2017, recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, che suggerisce «l'istituzione di uno status giuridico specifico per i robot nel lungo termine, di modo che almeno i robot autonomi più sofisticati possano essere considerati come persone elettroniche responsabili di risarcire qualsiasi danno da loro causato, nonché eventualmente il riconoscimento della personalità elettronica dei robot che prendono decisioni autonome o che interagiscono in modo indipendente con terzi»¹³. Pochi mesi dopo la risoluzione del Parlamento Europeo, che pone la necessità di ripensare la tradizionale teoria generale del diritto, il Comitato Nazionale per la Bioetica in Italia ha osservato 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)»¹⁴.

3. Pensiero "ibrido"

La possibilità di attribuire una vera e propria personalità elettronica a una macchina particolarmente intelligente, che presenta forma materiale (*hardware*) o immateriale (*software*) e che agisce anche senza controllo umano, problematizza il tema giusfilosofico del paragone tra

¹² 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* 12 (2010) 497-561; G. SARTOR, *Gli agenti software: nuovi soggetti del ciberdiritto?*, in *Contratto e impresa* 18 (2002) 465-499.

¹³ *Gazzetta Ufficiale dell'Unione Europea*, 18 luglio 2018, 2018/C 252/25, 239-257.

¹⁴ 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].

uomo e macchina e suggerisce di evitare di applicare a un robot sofisticato un'idea antropocentrica, che è tipica della visione individualistica della postmodernità, valutando attentamente la rilevanza giuridica del passaggio di un'entità robotica da *res* a *persona*¹⁵. Infatti il pensiero tecnologico dell'era contemporanea non può trascurare questa nuova "soggettività robotica", ma tende a presupporre che anche la macchina intelligente (come l'uomo) debba essere dotata primariamente di facoltà psicologiche e in questo paragone indebito omette di considerare la condizione biologica della corporeità umana¹⁶ oppure le interazioni sociali degli agenti *software*¹⁷, dimenticando che la condizione biologica e la relazionalità sono tratti originari della soggettività umana¹⁸.

È un dato di fatto che la macchina dotata di intelligenza artificiale avanzata operi ormai attivamente nell'odierna "infosfera", nella quale il tutto viene prima dell'agente e l'uomo non è l'unico soggetto cui attribuire una dignità morale, diventando così vero e proprio "organismo informazionale" (*infor*g), le cui condotte, intenzionali o meno, possono produrre effetti sulle entità interconnesse¹⁹.

I sistemi di intelligenza artificiale sono dotati da tempo non solo dell'abilità di calcolare, ossia di elaborare dati in base al modello computazionale dell'informatica, ma anche della capacità di svolgere azioni simili alle decisioni, ossia di eseguire comportamenti finalizzati in base a modelli di origine matematica, come quello cibernetico della retroazione (*feedback*) e quello informatico dell'apprendimento automatico (*machine*

¹⁵ U. RUFFOLO, *Il problema della "personalità elettronica"*, in *Journal of Ethics and Legal Technologies* 2 (1/2020) 75-88.

¹⁶ P. BENANTI, *The Cyborg: corpo e corporeità nell'epoca del post-umano. Prospettive antropologiche e riflessioni etiche per un discernimento morale*. Cittadella Editrice, Assisi 2012.

¹⁷ G. TEUBNER, *Die anonyme Matrix: Zu Menschenrechtsverletzungen durch „private“ transnationale Akteur*, in *Der Staat* 45 (2/2003) 161-187.

¹⁸ G. DE BONA, *Verso la tutela giuridica dei sistemi intelligenti. Prospettive critiche della soggettività robotica*, in *Journal of Ethics and Legal Technologies*, 4 (2/2022) 51-104.

¹⁹ L. FLORIDI, *La quarta Rivoluzione*, Raffaello Cortina, Milano 2017.

learning)²⁰. Soprattutto nel suo approccio odierno, però, l'intelligenza artificiale ottiene un'acquisizione di cognizioni e condotte che restano indipendenti dalla programmazione informatica e quindi dall'intervento umano²¹, sia che si tratti di algoritmo "supervisionato" – ossia quando il sistema (per esempio, la rete neurale) apprende il nesso che unisce dati particolari e ipotesi generali, imparando a classificare esempi simili – sia che si tratti di algoritmo "per rinforzo", ossia quando il sistema interagisce con l'ambiente ed esegue gli esempi di comportamenti più vicini e simili al risultato, che quindi migliorano l'azione da eseguire, evitando quelli più distanti e diversi, che sono considerati errati.

Costruendo automaticamente un modello dall'analisi dei dati su cui è addestrato, il sistema intelligente genera valutazioni e previsioni sui nuovi casi ad esso sottoposti in base al continuo aggiornamento dei dati medesimi, migliorando il modello predittivo attraverso l'ottimizzazione stocastica e la massimizzazione induttiva di una certa quantità probabilistica²².

Sebbene sia stato constatato che l'apprendimento automatico delle macchine possa puntare a raggiungere l'autentico "algoritmo dominante"²³, gli esiti del modello creato dall'intelligenza artificiale e le ragioni che ne giustificano le decisioni non possono essere spiegati dal codice sorgente o dai risultati osservabili della procedura impostata, giacché il funzionamento interno del sistema resta oscuro agli stessi programmatori e può essere considerato una *black box*²⁴. Peraltro, come si è già osservato, la capacità di apprendimento del

²⁰ G. TAMBURRINI, *I matematici e le macchine intelligenti. Spiegazione e unificazione nella scienza cognitiva*, Bruno Mondadori, Milano 2002.

²¹ S. RUSSELL - P. NORVIG, *Artificial Intelligence. A Modern Approach*, Prentice Hall, Englewood Cliffs (N. J.) 2010.

²² N. CRISTIANINI, *On the current paradigm in artificial intelligence*, in *AI Communications* 27 (2011) 37-43.

²³ P. DOMINGOS, *The Master Algorithm. How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, New York 2015.

²⁴ F. PASQUALE, *The black box society: The secret algorithms that control money and information*, Harvard University Press, Cambridge 2015.

robot evoluto è interattiva (*interactive learning*) e incrementa progressivamente la propria autonomia non soltanto apprendendo dall'ambiente e dagli oggetti che incontra, ma anche cooperando con l'essere umano, come avviene nella robotica collaborativa²⁵. L'impatto della regolazione algoritmica della società (per esempio, attraverso attività di *crowdsourcing* o di gestione della reputazione) ha anche suggerito la nozione di "macchina sociale", che genera evidenti conseguenze etiche e giuridiche per la tutela dell'autonomia individuale e dell'ordine pubblico²⁶. Pertanto assume peculiare interesse per l'etica e per il diritto la possibilità di ipotizzare che la macchina intelligente, presuntivamente dotata di facoltà artificiali analoghe, se non superiori, a quelle umane, possa essere titolare di una specie di autodeterminazione individuale e, quindi, di una sorta di libero arbitrio, come ha sostenuto John McCarthy²⁷. Infatti, se tale qualificazione fosse accettabile, occorrerebbe rivedere l'idea antropocentrica dell'intelligenza artificiale e prendere atto che l'agente robotico particolarmente sofisticato potrebbe acquisire una certa forma di libertà soggettiva e quindi, una pur imprecisata titolarità di diritti e di doveri²⁸.

È da notare che la programmazione di decisioni consapevoli e l'applicazione di regole condivise possono essere imposte al sistema robotico da algoritmi di istruzione e di apprendimento, ma tale questione supera i confini di una forma controllabile di intelligenza artificiale²⁹. Difatti,

²⁵ M. A. GOODRICH - A. C. SCHULTZ, *Human-Robot Interaction: A Survey*, in *Foundations and Trends in Human-Computer Interaction*, 1 (3/2007) 203-275.

²⁶ 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>].

²⁷ J. MCCARTHY, *Free Will - even for Robots*, in *Journal of Experimental and Theoretical Artificial Intelligence* 12 (3/2000) 341-352.

²⁸ 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.

²⁹ W. WALLACH - C. ALLEN, *Moral machines. Teaching robots right from wrong*, Oxford University Press, Oxford 2009.

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 e incertezza e sono difficili da codificare attraverso misurazioni o funzioni numeriche³⁰.

Il paragone tra uomo e macchina costituisce il presupposto apodittico della metafora "intelligenza artificiale" e appare ipotizzato da molti studiosi nelle scienze naturali, morali e sociali, che di regola sostengono prospettive riferibili all'odierno postumanesimo e all'intelligenza artificiale cd. "forte"³¹. La recente evoluzione negli studi e nelle applicazioni dell'intelligenza artificiale e della biorobotica ha portato a considerare l'avvento della "singolarità" come un evento probabile. Con questo termine si fa riferimento al momento storico in cui l'intelligenza delle macchine informatiche e dei robot più avanzati diventerebbe tale da superare le capacità umane, generando un cambiamento epocale paragonabile all'emergere della vita sulla Terra³². Questo punto critico dell'evoluzione consentirebbe un'interazione più stretta tra la mente umana e l'intelligenza artificiale e questo traguardo sarebbe raggiunto nel breve periodo grazie alla legge dei "ritorni acceleranti", secondo la quale la tecnologia progredisce in modo esponenziale, superando rapidamente le capacità del pensiero umano³³.

Pur essendo tuttora incerto se la singolarità tecnologica possa o debba realmente giungere a compimento, questa situazione non distingue l'intelligenza meccanica da quella umana, ma ne prefigura una commistione biorobotica, introducendo l'era del pensiero ibrido (*hybrid*

³⁰ Z. LIPTON, *The mythos of model interpretability*, in *ICML Workshop on Human Interpretability*, 2016, arXiv:1511.03677.

³¹ R. PEPPERED, *The Posthuman Manifesto*, in *Kritikos 2* (2005) 1-16.

³² V. VINGE, *The Coming Technological Singularity: How to Survive in the Post-Human Era*, in *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace*, Proceedings of a symposium cosponsored by the NASA Lewis Research Center and the Ohio Aerospace Institute and held in Westlake, Ohio 1993, 11-22.

³³ R. KURZWEIL, *The singularity is near. When humans transcend biology*, Viking, New York 2005.

thinking)³⁴. Ne conseguirebbe l'opportunità di riflettere criticamente sulle conseguenze etiche e giuridiche del sopravvento della singolarità, concepita non solo e non tanto come una tappa nell'evoluzione della specie umana oppure delle macchine superintelligenti, ma come una nuova ontologia in cui potrebbe disperdersi la distinzione tra il naturale e l'artificiale oppure tra il soggetto e l'oggetto.

4. Prospettiva umana

Nell'indirizzo filosofico che ammette la nozione di intelligenza artificiale "debole" si accoglie criticamente una prospettiva ibrida e si accetta l'inevitabile raffronto tra uomo e *robot*, ma resta giustamente dubbia la reale analogia tra struttura antropologica e macchina intelligente. Infatti, lo sforzo di innalzare il dispositivo tecnologico dotato di intelligenza artificiale al livello delle facoltà umane si realizza attraverso operazioni di formalizzazione che, in sostanza, si traducono nella costruzione di algoritmi informatici che governano e accelerano il presunto ragionamento e la conseguente azione del *robot*. Come si è già precedentemente illustrato, il paragone è infondato e si fonda sul razionalismo moderno, secondo il quale l'uomo sarebbe capace di conoscere ciò che è bene e ciò che è male attraverso l'uso della propria ragione, e ha contribuito a fondare i programmi di intelligenza artificiale sul metodo scientifico³⁵, riconducendo a schemi comuni di origine algoritmica le macchine e i sistemi biologici capaci di esibire un comportamento finalizzato, confermando il presupposto matematico della cibernetica³⁶.

³⁴ R. KURZWEIL, *Get ready for hybrid thinking*, 2014, TED Talk [https://www.ted.com/talks/ray_kurzweil_get_ready_for_hybrid_thinking?language=en]

³⁵ D.A. GILLIES, *Artificial intelligence and scientific method*, Oxford University Press, Oxford 1996.

³⁶ G. TAMBURRINI, *I matematici e le macchine intelligenti. Spiegazione e unificazione nella scienza cognitiva*, Mondadori, Milano 2002.

L'intelligenza umana non è dimostrabile con un'equazione: dunque, questo apparente innalzamento della tecnologia digitale, che forse sorpassa l'uomo in alcuni dei suoi aspetti, è un abbassamento della macchina intelligente a una parte delle capacità logiche del soggetto umano che, secondo la convinzione del pensiero computazionale, ragiona nel modo migliore e, quindi, è intelligente soltanto quando calcola o, comunque, quando esegue le istruzioni degli algoritmi di programmazione.

La coscienza rimane una proprietà irriducibile 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: infatti, la coscienza genera un problema che non è computabile³⁷. Dunque, l'equivalenza tra mente umana e cervello elettronico è inammissibile, perché le macchine calcolanti sono inconsapevoli. Infatti, applicando gli algoritmi di programmazione che ogni *computer* deve comunque eseguire, il dispositivo cibernetico elabora un ragionamento meccanico senza comprendere ciò che dice o ciò che fa. Tutte le logiche di tipo formale, che sono largamente utilizzate nelle macchine intelligenti, si manifestano tramite un lessico e una sintassi di tipo rigidamente simbolico, analizzabile seguendo regole di inferenza che hanno natura algoritmica, deduttiva o induttiva. La comprensione del linguaggio e l'interpretazione delle regole inferenziali da parte dell'operatore che ne fa uso è indipendente dallo sviluppo e dall'esito del procedimento inferenziale.

La mente umana non può essere un sistema di intelligenza artificiale, che 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

³⁷ F. FAGGIN, *Irriducibile. La coscienza, la vita. i computer e la nostra natura*, Mondadori, Milano 2022.

linguaggio e comprensione significativa delle operazioni combinatorie³⁸. La logica proposizionale booleana, spesso considerata un postulato del ragionamento informatizzato, non è in grado di realizzare un qualsiasi sistema complesso capace di superare la morfologia sintattica del linguaggio per giungere alla sua piena comprensione semantica. Peraltro, i modelli statistici non sono adatti a cogliere la struttura sintattica e innata del linguaggio³⁹, e dunque l'intelligenza artificiale (compresa quella generativa) non è in grado di generare automaticamente la grammatica delle lingue umane che, talvolta, seguono regole "impossibili"⁴⁰.

5. Conclusione

In conclusione, l'avvento della nuova legislazione europea, che disciplina il fenomeno con un approccio basato sul livello di rischio di dannosità derivante dall'uso di tale tecnologia, documenta l'importanza di definire correttamente il significato dell'espressione "intelligenza artificiale", riferendosi alla sua origine matematica e ai suoi sviluppi applicativi nell'informatica e nella robotica. Riflettendo criticamente sul paragone tra uomo e macchina intelligente, diventa impossibile stabilire un'analogia tra stato mentale e logica della macchina, che avviene solo sul piano dell'accettabilità dei risultati e non della identità delle procedure di ragionamento e di decisione: tali metodologie sono riferibili propriamente soltanto all'uomo e sono radicalmente differenti nel calcolatore, che tuttavia può arrivare ai medesimi esiti dimostrativi⁴¹. Ogni algoritmo che riguarda le applicazioni dell'intelligenza artificiale è frutto di una programmazione assiomatica e rimane opaco, sebbene

³⁸ J.R. SEARLE, *Minds, brains, and programs*, in *Behavioral and Brain Sciences* 3 (3/1980) 417-457.

³⁹ N. CHOMSKY, *Syntactic Structures*, Mouton, The Hague 1957.

⁴⁰ A. MORO, *Impossible languages*, MIT Press, Cambridge 2016.

⁴¹ S. CRAFA, *Artificial Intelligence and Human Dialogue*, in *Journal of Ethics and Legal Technologies* 1 (2019) 44-56.

influisca sull'obiettività dell'informazione e sulla riservatezza, condizionando argomentazioni ed emozioni che si propagano in rete: algoritmi che pretendono di interpretare e comprendere argomenti, opinioni, emozioni, restano ipotesi sottratte alla discussione e formulate da chi detiene il potere di elaborarle ed amministrarle.

Il probabile conflitto fra il libero arbitrio della comunità umana e l'autonomia dei sistemi di intelligenza artificiale, prefigurato nel secolo scorso dalla migliore letteratura fantascientifica⁴², non sembra lontano dalla realtà, ma richiede il ripensamento di un'etica della sostenibilità tecnologica che oltrepassi le frontiere del soggettivismo e consenta un'equilibrata convivenza con le macchine intelligenti⁴³. L'intelligenza artificiale sta realizzando trasformazioni economiche e sociali con un'accelerazione esponenziale e talmente invasiva da indurre il diritto europeo ad approntarne una meditata regolazione. Come dimostra la preparazione della proposta di regolamento dell'Unione Europea, l'esigenza di organizzare un codice giuridico dell'intelligenza artificiale con contenuti normativi innovativi e con un approccio pluralista, che tenga conto dei differenti valori ed interessi coinvolti, è l'autentica sfida posta dai problemi controversi dell'odierno tecnodiritto⁴⁴. Infatti, il cambiamento di paradigma che l'intelligenza artificiale sta imprimendo sulle relazioni sociali e sull'agire umano sta mutando anche il concetto e la funzione del diritto, sollevando questioni che non riguardano solo l'interpretazione di nuove norme positive o inesplorate decisioni della giurisprudenza, ma richiedono lo studio critico delle definizioni e dei principi di un fenomeno tecnologico tuttora in evoluzione.

⁴² P.K. DICK, *Do Androids Dream of Electric Sheep?*, Doubleday and Company, New York 1968.

⁴³ 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.

⁴⁴ P. MORO - C. SARRA (curr.), *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, FrancoAngeli, Milano 2017.

“LUOGHI ETICI” E COMPETENZE ETICHE NELL'EPOCA DELL'INTELLIGENZA ARTIFICIALE

Leopoldo Sandonà

Abstract

La presenza di “luoghi etici” legati all'Intelligenza Artificiale è ancora relativamente scarsa. Confrontando alcune esperienze con la tradizione dei Comitati etici in ambito clinico si può giungere ad alcune interessanti conclusioni sul rapporto tra *expertise* morale e competenza etica organizzativa.

Parole chiave: IA, Bioetica, “Luoghi etici”, Competenza etica, Organismi regolatori.

The presence of ethics committees related to Artificial Intelligence is still relatively scarce. Comparing some experiences with the tradition of ethics committees in clinical field, we can foster interesting conclusions about the relationship between moral expertise and organisational ethical competence.

Keywords: AI, Bioethics, “Ethical places”, Moral expertise, Regulatory bodies.

Nessuno ci obbliga a considerare come inevitabile il progresso tecnologico quando questo non è utile al perseguimento delle finalità dell'uomo. La libertà vera è quella che consente alle diverse comunità politiche di proteggere i valori che considerano importanti, ed è questa la libertà che oggi dobbiamo esercitare¹.

¹ F. FUKUYAMA, *L'uomo oltre l'uomo. Le conseguenze della rivoluzione biotecnologica*, Mondadori, Milano 2003, 298.

1. Premesse

Anche se le prospettive di Fukuyama non si inseriscono nel quadro della riflessione sull’Intelligenza Artificiale (AI), alcune delle conseguenze riferite alla rivoluzione biotecnologica di qualche decennio fa sono applicabili anche alla situazione attuale. In questa direzione non è superfluo affrontare il nostro tema dal punto di vista delle ricadute etico-pratiche, ma soprattutto dal punto di vista di quei “luoghi” che possono prevenire e affrontare i problemi etici connessi con l’AI. Nell’accostarci a un tema così delicato e nuovo, sono necessarie anzitutto alcune premesse di ordine contestuale: siamo dentro un quadro in evoluzione, che sarebbe assai erroneo fotografare secondo istantanee ferme. Il compito realistico è dunque quello di cercare di “balbettare” qualcosa in questo vorticoso sviluppo. Quando mi è stato proposto questo intervento, vista la presenza di molti interventi etico-antropologici, ho deciso di concentrarmi sull’aspetto specifico dei “luoghi etici” di riferimento. Una seconda premessa è legata esplicitamente al tema: come afferma Adriano Pessina,

«se tutta la realtà è rappresentata sulla base di *modelli* e si accetta [...] di sostituire l’autocomprensione dell’umano, giudicata inattendibile per qualsiasi ricerca scientifica, con un *modello sperimentale di tipo informazionale*, si finisce con il *credere* che tutto quanto non rientri nel modello sia insignificante»².

Tale prospettiva determina uno snodo contraddittorio, perché rischiamo di aver dato vita a un modello che diviene ideale regolativo, modellando il modellatore, ricreando cioè il creatore dello stesso modello. Oltre tale modello si staglia il nulla. Come facciamo a parlare di deliberazione etica, di argomentazione etica e di “luoghi etici” se siamo già in un modello predeciso e precostituito che si autogenera prescindendo ormai dal suo programmatore? Lo snodo tragico sta nel fatto che tali modelli, definiti dall’uomo, tendono oramai ad escluderlo

² A. PESSINA, *L’essere altrove. L’esperienza umana nell’epoca dell’intelligenza artificiale*, Mimesis, Milano-Udine 2023, 87 [corsivi dell’Autore].

nel loro sviluppo più radicale. In un contesto in cui il *vero* è ridotto all’*esatto* – e potremmo dire anche che il *giusto* è ridotto all’*esatto* – sembra non esserci più spazio per luoghi di deliberazione etica, in una società senza rito, senza dolore, senza carne³, che vive una radicale tendenza narcisistica e gnostica⁴. Nell’accelerazione del sistema tecnico⁵, l’anima-
le simbolico svanisce⁶. Se il modello esclude l’uomo è quasi impossibile parlare di deliberazione etica.

L’ultima premessa riguarda l’impatto dell’etica: come ricorda Luciano Floridi⁷, siamo dentro questa infosfera e non possiamo restaurare la casa in cui siamo se non dall’interno, dobbiamo riparare la barca stando all’interno e non portandola comodamente in un cantiere, non possiamo scendere dalla bicicletta e portarla in officina ma dobbiamo pedalare e insieme determinare un cambiamento dall’interno e non in forma estrinseca.

2. Per un confronto con i “luoghi etici” in ambito medico-clinico

Come primo passaggio appare necessario e fondamentale un confronto con l’esperienza dei “luoghi etici” tradizionali in ambito medico-clinico, in particolare i Comitati etici per la pratica clinica⁸, luoghi

³ Cf. B.C. HAN, *La società senza dolore: perché abbiamo bandito la sofferenza dalle nostre vite*, Einaudi, Torino 2022; ID., *La scomparsa dei riti: una topologia del presente*, Nottetempo, Milano 2021; ID., *Le non cose: come abbiamo smesso di vivere il reale*, Einaudi, Torino 2022.

⁴ Cf. V. POSSENTI, *IA. La pericolosa equivocità del termine*, *supra*; PESSINA, *L’essere altrove*, cit., 119-123.

⁵ Cf. J. ELLUL, *Il sistema tecnico: la gabbia delle società contemporanee*, Jaca Book, Milano 2009.

⁶ Cf. E. CASSIRER, *Filosofia delle forme simboliche*, Pgreco, Milano 2015.

⁷ L. FLORIDI, *La quarta rivoluzione: come l’infosfera sta trasformando il mondo*, Raffaello Cortina, Milano 2017, 224.

⁸ Nel nostro Paese sono presenti i Comitati etici per la sperimentazione clinica (Cesc), con natura deliberativa, recentemente definiti come Comitati etici territoriali (Cet) dal DM del Ministero della Salute del 26 gennaio 2023. Solo in poche regioni (come Triveneto e Toscana) e in alcuni policlinici sono presenti i Comitati etici per la pratica clinica (Cepc),

dell’etica consultivi che la filosofia morale ha condiviso in questi decenni con altre discipline, in particolare medico-cliniche ma anche delle scienze sociali. A partire da questa analisi e dal confronto con le poche esperienze di Comitati etici per l’Intelligenza Artificiale, si può riconoscere una centralità dei regolamenti di tipo economico-giuridico⁹, centralità che non va demonizzata ma precisata nelle sue forme e nei suoi sviluppi. Il rischio peraltro in questo contesto è che l’etica venga strumentalizzata per finalità altre. Confrontandoci con l’ambito AI, la domanda da porsi è: a quale traguardo sono giunti i Comitati etici in ambito clinico? Non siamo, ad essere sinceri, in un grande momento. Sia Cesc che Cepc, negli Stati Uniti d’America e poi in Europa, sono nati prima della nascita ufficiale della bioetica avvenuta negli anni Settanta con Van Rensselaer Potter. Come sta accadendo per l’AI, prima sono nati i problemi e poi si è pensato a limitare tali derive. La differenza è che ora l’AI ha la possibilità di affrontare i problemi non *ex post* ma *ex ante* e *in itinere* all’interno delle sperimentazioni, attraverso un accompagnamento continuo.

L’insoddisfazione per i sentieri della bioetica, più volte denunciata¹⁰ a più di cinquant’anni dalla nascita della disciplina, è divenuta negli ultimi anni anche insoddisfazione per i “luoghi etici”, da un lato divenuti (nei Cesc) luoghi di standardizzazione, dall’altro ridotti nei Cepc ad una bioetica “difensiva” nei casi specie di fine vita, in cui i Comitati

con natura consultiva. Più volte il Comitato nazionale per la bioetica, fino al documento *I Comitati per l’etica nella clinica* (31 marzo 2017), ha posto in evidenza la necessità di giungere a una configurazione nazionale di questi ultimi Comitati, per la loro preziosa funzione di prossimità nella pratica clinica.

⁹ Anche nell’AI Act del 14 giugno 2023, con un’ulteriore approvazione del 9 dicembre 2023, la centralità economica è prevalente. Accanto a questo documento è doveroso ricordare il documento del 2020 *The Assessment List for Trustworthy Artificial Intelligence (Altai) for Self Assessment*, anche per il riferimento non scontato a un’AI *trustworthy* (affidabile/degna di fiducia). La strada per l’adozione di queste misure a livello operativo è ancora lunga, ma il percorso europeo, non privo di limiti, sembra rappresentare un’incoraggiante terza via tra le prospettive nordamericane e l’utilizzo del regime cinese.

¹⁰ A. PESSINA, *Bioetica: l’uomo sperimentale*, Pearson, Milano-Torino 2020, 33-37.

vengono chiamati in causa dopo una prospettiva multidisciplinare per ratificare un percorso già predefinito¹¹. Siamo giunti non a una bioetica che accompagna il caso, che lo valuta strada facendo¹², ma a una bioetica cui non resta che dare una ratifica finale. Questa insoddisfazione non sembra essere l’ultima parola se, come avvenuto negli ultimi dieci anni, ha contribuito a formare vie alternative, dalla consulenza etica¹³ alla crescita degli spazi etici, specie nella tradizione francofona¹⁴.

La traiettoria di una bioetica difensiva rischia di riproporsi, con molti addentellati nella *business ethics* e nell’etica delle organizzazioni, anche nel versante AI, all’interno di un certo *ethical washing* o *window dressing*¹⁵, espressioni che stanno ad indicare un utilizzo dell’etica strumentale e reputazionale per “lavare” qualche “peccato” nell’operatività ordinaria.

¹¹ CONSIGLIO REGIONALE DEL VENETO, *Progetto di Legge n. 217 di iniziativa popolare*, art. 3. Il Comitato etico, dopo la valutazione della Commissione tecnica multidisciplinare, si trova nella condizione di asseverare un processo già definito in precedenza, con una pressione mediatica e organizzativa non indifferente che lede l’autonomia su cui si fonda la natura degli stessi Comitati.

¹² Così voleva la normativa che regolava le funzioni dei Comitati etici per la pratica clinica: REGIONE DEL VENETO, *Disciplina della rete dei Comitati etici: riordino delle disposizioni relative al Comitato Regionale per la Bioetica e ai Comitati Etici per la Pratica Clinica*, DGR n. 983/2014, all. B, art. 5.1.a.: «1. L’analisi etica di casi clinici ha la finalità di promuovere decisioni eticamente giustificate, per garantire il rispetto del paziente e supportare gli operatori sanitari e l’istituzione sanitaria nella scelta del percorso più corretto eticamente, avvalendosi, ove ritenuto necessario, della consulenza di esperti esterni in considerazione dello specifico caso clinico da esaminare. 2. Il parere del Comitato costituisce un supporto per chiarire le questioni etiche connesse a una storia clinica, per approfondire le problematiche etiche associate alle decisioni sulle cure individuali e prevenire possibili conflitti nella gestione dello stato di salute dei pazienti».

¹³ GRUPPO NAZIONALE DI ETICA CLINICA E CONSULENZA ETICA IN AMBITO SANITARIO, *Documento di Trento. La consulenza etica in ambito sanitario in Italia*, 10 ottobre 2013.

¹⁴ COMITATO NAZIONALE PER LA BIOETICA, *Vulnerabilità e cura nel welfare di comunità. Il ruolo dello Spazio etico per un dibattito pubblico*, 10 dicembre 2021.

¹⁵ Parallelamente alla prospettiva del *greenwashing* etico per l’ambiente, si parla di *bluwashing* etico per l’ambito digitale; Cf. L. FLORIDI, *Etica dell’intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina, Milano 2022, 111.

3. Esperienze di “luoghi etici” nell’ambito AI

La ricerca di una riflessione etica sull’AI 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 leggibili¹⁶.

Un altro ambito che non prenderemo in considerazione sono i codici deontologici e di condotta. In questa direzione si muove in fondo anche il Regolamento europeo, che non cita in modo intensivo la dimensione etica. Tali codici definiscono una dinamica consultiva e disciplinare, o addirittura aspirazionale/motivazionale. Sia nel caso dei codici etici che dei codici deontologici o comportamentali non si vuole negare una loro specifica importanza, ma essi vanno inseriti in un’assunzione della tematica etica non occasionale né strumentale, bensì strutturale.

A questo elemento, seppur sempre in un’ottica di utilità aziendale, sembra invece fare riferimento una strutturazione sistematica di Comitato etico per l’AI¹⁷. A partire dagli errori dell’AI (ad es. nell’addestramento di essa solo con medici maschi e occidentali, provocando così delle forme discriminatorie dell’AI per altri gruppi di cittadini) i rischi etici nell’AI sono sempre più evidenti: «pregiudizi involontari e violazioni della privacy sono solo due dei più visibili»¹⁸. Ci sono elementi che non possono essere limitati semplicemente con una dimensione tecnica e per questo è necessario un Comitato, con la funzione di «identificare sistematicamente ed esaurientemente i rischi etici di prodotto

¹⁶ *Ibid.*, 91 ss., ove si parla esplicitamente di cinque principi AI in cui, ai quattro classici, si aggiunge il quinto dell’esplicabilità.

¹⁷ <https://www.hbritalia.it/settembre-2022/2022/09/01/pdf/perche-vi-occorre-un-comitato-etico-per-lia-15328/?id=0> (accesso 30.11.2023).

¹⁸ *Ibid.*

basati sull’AI che vengono sviluppati [...] e contribuire ad attenuarli»¹⁹. Al Comitato, dunque, non basterebbe una dimensione meramente consultiva, in quanto dovrebbe avere diritto di veto e di intervento nelle scelte aziendali. Un aspetto molto rilevante è l’azione di fiducia che il Comitato riesce a instaurare non solo nella dimensione intra-aziendale ma anche con i clienti esterni, i consumatori e gli *stakeholders*. I membri dovrebbero essere umanisti di area filosofica, avvocati, esperti di strategia aziendale, tecnologi e persone in grado di individuare i *bias* del sistema già nelle prime fasi del processo. Gli esperti di etica, in particolare, farebbero parte del Comitato

«perché hanno la formazione, le conoscenze e le esperienze che occorrono per identificare e soppesare una vasta gamma di rischi etici, hanno familiarità con concetti e distinzioni che sono preziosi nelle scelte etiche e possono aiutare i vari gruppi a valutare obiettivamente i problemi etici»²⁰.

Tali esperti avrebbero anche una funzione formativa sulle tematiche specifiche, in modo da allargare e radicare la cultura aziendale in ambito etico. Se è vero, dunque, che questo documento si pone in un’ottica difensiva e di tipo reputazionale²¹, rendendo il Comitato anzitutto conveniente, la centralità di quest’organo/ufficio appare importante, prevenendo i rischi etici per l’azienda. A partire da questo esempio, i vantaggi dell’etica nell’AI rispetto all’ambito medico sanitario si riassumono nella possibilità di accompagnare il processo, ma ci si può ricollegare anche alle funzioni tradizionalmente presenti nei Cepc: analisi di casi, formazione, elaborazione di linee di indirizzo e consulenza nell’allocazione delle risorse²². Nella formazione emerge non solo un

¹⁹ *Ibid.*

²⁰ *Ibid.*

²¹ *Ibid.*: «non dedicare la massima attenzione ai criteri di formazione del comitato e al ruolo che dovrà assumere all’interno della vostra organizzazione potrebbe essere devastante per la reputazione della vostra azienda».

²² Tale funzione, che appare essere sostanzialmente non sviluppata, è assai significativa non solo rispetto all’attualità del tema allocativo, ma anche in rapporto alla stretta interconnessione tra etica clinica e organizzativa, questione sempre più urgente in ambito

riferimento alla formazione degli specialisti, ma anche un’informazione dell’opinione pubblica che esalti la prospettiva di un pensiero critico.

Come detto, il vantaggio principale rispetto al mondo sanitario è poter seguire *in itinere* ed *ex ante* lo sviluppo delle tecnologie, potendo avere anche dall’utente un report su ciò che funziona e su ciò che non funziona. In questa prospettiva non ci si muove né con ottimismo né con pessimismo, ma entro un quadro critico e sapienziale che impone adeguati «assalti di lucidità»²³. Abbiamo imparato dalla storia della bioetica che essa non si è radicata nelle istituzioni e così facendo ha maturato una fragilità intrinseca, a cavallo tra medicina legale e filosofia morale. Qui invece l’etica può accompagnare il processo “a monte” e non “a valle”, come fa invece ormai una certa casistica biogiuridica. Si può recuperare il significato delle cellule del buon consiglio pre- e post- istituzionali di cui parla Paul Ricoeur²⁴. Inoltre, rispetto al modello fisso che finisce per determinare degli standard non-umani, *l’ethics by design* consente di definire un continuo accompagnamento del processo di programmazione, aperto e mai chiuso²⁵.

4. Abilità etiche, moral expertise e competenza etica

A partire dallo scenario qui definito, appare fondamentale l’“addestramento etico” delle nuove tecnologie. Se una strada è quella di fare riferimento alle cosiddette *soft skills* o “abilità spontanee” – con gli antichi potremmo chiamarle “virtù naturali” – 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

socio-sanitario e rilanciata prepotentemente dallo sviluppo dell’AI.

²³ Cf. G. GOISIS, *La sfida del “tecnoumano”, infra.*

²⁴ P. RICOEUR, *Etica e vivere bene: conversazione con Paul Ricoeur*, in AA.VV., *Il male*, Raffaello Cortina, Milano 2000, 6.

²⁵ V. PAGLIA - R. PEGORARO (curr.), *The Good Algorithm? Artificial Intelligence: Ethics, Law, Health*, Pontifical Academy for Life-Libreria Editrice Vaticana, Città del Vaticano 2020.

moral expertise ed *ethical competence* sembrano allargare il compito, 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 ad 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 condivisa. Tutti abbiamo problemi morali e tutti avremo problemi morali in ambito di AI. Anche se non tutti saremo filosofi morali è fondamentale che tale competenza venga fatta crescere e ci abitui a riconoscere determinate problematiche.

5. Conclusioni aperte

Concludendo questo intervento appare chiaro che non ci sono risposte “chiare e distinte” per un tema complesso e in vorticoso evoluzione. È anzitutto fondamentale evitare di vivere questa evoluzione come un progresso solo quantitativo che si autoregola. È invece importante che i singoli cittadini si sentano responsabili dell’implementazione qualitativa

di tale sviluppo e per fare ciò è assolutamente prioritario il pensiero critico come condivisione sociale di prospettive etiche. In chiave specifica la filosofia avrà un ruolo decisivo²⁶ per capire chi allena questa macchina, lasciando lo spazio aperto per non incorrere nei pregiudizi, spesso impliciti e inavvertiti, di chi plasma l’AI. Proprio come nella caverna platonica, possiamo illuderci di vivere di una verità che però si rivela essere solo un’ombra del vero mondo. La filosofia, pensiero sempre antico e sempre nuovo, ci può consentire una liberazione dalle illusioni e dalla falsa comprensione del mondo. In questi «assalti di lucidità» potrebbe molto ispirare la prospettiva di Romano Guardini, che indica la strada dell’*ascesi del potere* e non *dal potere*, in cui il potere si pone come verbo e come possibilità aperta, in cui lo stesso potere ha in se stesso la radice della sua autolimitazione²⁷.

Sul piano operativo, esattamente come il *Belmont Report* è stato un punto di partenza fondamentale per la bioetica degli anni Settanta, così il Regolamento appena varato dall’Unione Europea è un punto di partenza per la riflessione sull’impatto dell’AI. Da quel regolamento bisogna partire, senza tacerne alcuni limiti, per istruire uno sviluppo anche a livello etico. Non possiamo negare che questa AI ci porta non solo a verificare un grande progresso ma anche la vulnerabilità del sistema, un mondo che ci fa vedere da un lato grandi sviluppi e dall’altro serve all’uomo per tornare nella barbarie (una barbarie ipertecnologica). Per fare un esempio, il drone porterà direttamente a casa i nostri acquisti ma insieme il drone dà la morte dal cielo come recentemente è accaduto nella riapertura del conflitto israelo-palestinese. Il tema della vulnerabilità è centrale facendo attenzione alla vulnerabilità che si crea. Essa da un lato ci fa sentire limitati, come cavalli superati dalle macchine, ma

²⁶ Cf. M. KRIENKE, *Quali sono le sfide dell’IA per la libertà? Implicazioni etico-politiche*, cf. *supra*; A. GALLUZZI, *Tecnologie del sé e il caso dei gemelli digitali*, *supra*.

²⁷ R. GUARDINI, *La fine dell’epoca moderna. Il potere*, Morcelliana, Brescia 2022. Appare utile notare come Guardini avesse inizialmente elaborato questo intervento preparando dei corsi su Pascal, visto dall’autore italo-tedesco come espressione di un’altra e più equilibrata modernità rispetto al versante “cartesiano”.

appare anche essere indice delle ingiustizie, non solo come luogo antropologico²⁸. I luoghi etici sono quella dimensione necessaria che può attenuare i rischi dell’applicazione ma che sa anche verificare quando le tecnologie vanno fuori strada.

Concludo con due immagini. La prima è la descrizione della vicenda dell’apostolo Paolo, che ha sfruttato le “autostrade” romane dell’epoca per la sua missione in Asia Minore²⁹. Non credo che Paolo condividesse tutto dell’Impero che stava perseguitando i cristiani, come lui ben sapeva. Il messaggio che traggo è che non dobbiamo demonizzare le infrastrutture in cui siamo dentro ma dobbiamo abitarle, starci dentro e camminarci: l’unificazione planetaria prodotta da queste tecnologie, se non sfocia nella massificazione, può avere molti esiti interessanti. La seconda immagine – teologica anche se proveniente da un filosofo morale – è tratta dal già citato testo di Adriano Pessina, che mette l’accento sul rischio di deriva gnostica e disincarnata di una certa impostazione tecnologica:

«se la storia dell’Occidente può essere lenta lungo questi crinali, l’affermazione della Presenza della Trascendenza nella storia, che è tesi propria dell’Incarnazione, si offre, però, come riconciliazione tra l’altrove e il qui e ora, cambiando la prospettiva: il senso ultimo dell’esistere e dell’essere non è altrove e il qui e ora non è solo la prigione storica dell’uomo»³⁰.

Senza essere un pulpito moralistico, i “luoghi etici” anche nell’ambito dell’AI possono essere una piccola testimonianza di reincarnazione dell’umano nella cura del sé che diviene cura delle strutture tecnologiche affidateci, un nuovo modo di declinare l’antico “conosci te stesso” e la biblica custodia del giardino in cui viviamo.

²⁸ Cf. H. TEN HAVE, *Vulnerability. Challenging Bioethics*, Routledge, London-New York 2016.

²⁹ J. MURPHY-O’CONNOR, *Paolo: un uomo inquieto, un apostolo instancabile*, San Paolo, Cinisello Balsamo (MI) 2008.

³⁰ A. PESSINA, *L’essere altrove. L’esperienza umana nell’epoca dell’intelligenza artificiale*, cit., 121.

Appendice

Dopo aver svolto la relazione e come promesso nella stessa ho sottoposto il titolo del mio intervento a OpenAI-Chat GPT. Ecco il risultato ottenuto.

La discussione sui luoghi etici e sulle competenze morali nell’ambito dell’AI è di fondamentale importanza, poiché l’AI sta diventando sempre più parte integrante della nostra vita quotidiana e sta influenzando una vasta gamma di settori, tra cui la salute, l’automazione industriale, l’assistenza all’anziano, la guida autonoma, il settore finanziario e molto altro. Qui di seguito esploreremo alcune considerazioni chiave:

1. “Luoghi etici” in cui l’AI viene utilizzata:

- *Assistenza sanitaria*: l’AI viene utilizzata per la diagnosi e il trattamento di malattie, ma è fondamentale che i medici e i ricercatori definiscano le linee guida etiche per l’uso dell’AI nella cura dei pazienti, il rispetto della privacy dei dati sanitari e l’equità nell’accesso ai servizi di salute.

- *Automazione industriale*: l’AI è utilizzata in settori come la produzione, la logistica e la gestione delle risorse umane. Qui è importante garantire che l’automazione rispetti i diritti dei lavoratori, le norme di sicurezza e l’impatto sull’ambiente.

- *Veicoli autonomi*: l’AI è una componente chiave dei veicoli autonomi. In questo contesto, l’etica è legata alla sicurezza stradale, alla responsabilità legale e alle decisioni che l’AI deve prendere in situazioni di emergenza.

- *Sistemi giudiziari e applicazione della legge*: l’AI è utilizzata per il monitoraggio e il controllo in ambienti come le prigioni o per l’analisi dei dati di sorveglianza. Qui è importante evitare discriminazioni e abusi di potere.

2. Competenze morali nell’AI:

- *Trasparenza e spiegabilità*: Gli sviluppatori di AI dovrebbero garantire che i sistemi siano trasparenti e spiegabili, in modo che le decisioni

prese dall’AI possano essere comprese e analizzate da esperti e dagli utenti stessi.

- *Equità*: L’AI non dovrebbe discriminare in base a razza, genere, religione o altre caratteristiche personali. I modelli di AI dovrebbero essere addestrati su dati rappresentativi e bilanciati.

- *Privacy*: Gli sviluppatori di AI devono proteggere la privacy dei dati degli utenti e garantire che i dati sensibili siano gestiti in modo sicuro.

- *Responsabilità*: Le organizzazioni che sviluppano e utilizzano l’AI devono assumersi la responsabilità delle azioni dei sistemi AI e delle conseguenze delle loro decisioni.

- *Sviluppo sostenibile*: L’AI dovrebbe essere utilizzata per sostenere gli obiettivi di sviluppo sostenibile, riducendo al minimo l’impatto ambientale e promuovendo una società equa.

- *Etica della robotica*: L’AI in robotica deve rispettare principi etici, come l’obbligo di evitare il danno agli esseri umani, di rispettare la dignità umana e di agire per il beneficio di tutti.

3. Organismi regolatori ed etica dell’IA:

Molti paesi stanno sviluppando linee guida etiche e normative per l’uso dell’IA. Questi organismi regolatori dovrebbero essere responsabili di garantire che l’IA sia utilizzata in modo etico e che vengano applicate sanzioni per il mancato rispetto delle regole.

4. Educazione e formazione:

Gli sviluppatori e gli utenti di AI dovrebbero essere formati in etica dell’IA per comprendere le sfide etiche e le competenze morali necessarie per sviluppare e utilizzare l’IA in modo responsabile.

L’etica e le competenze morali nell’IA sono cruciali per garantire che questa tecnologia contribuisca positivamente alla società e non porti a conseguenze indesiderate. È un campo in rapida evoluzione in cui vengono discussi continuamente nuovi problemi etici e sviluppate nuove competenze per affrontarli.

LA SFIDA DEL "TECNOUMANO"

Giuseppe Goisis

Fra le più grandi scoperte che cambiano decisamente il volto della nostra autoconsapevolezza, si allineano quelle arrecate dalle *neuroscienze*: con singolare rilevanza gli studi sul cervello umano hanno condotto a modificare sostanzialmente l'immagine che noi avevamo di noi stessi e delle nostre iniziative. Molto di ciò che la tradizione metafisica chiamava "spirito", molti suoi privilegi devono oggi essere rimodulati e attribuiti alla straordinaria *potenza* del nostro cervello, e soprattutto alla sua *plasticità*: pur così precario, esso può contenere l'idea di infinito, e anche se certamente non può che conoscere una piccola parte degli enigmi che si squadernano nel mondo, può immaginare, con l'aiuto della fantasia, tale molteplicità e creare universi, per dir così, paralleli¹. L'uomo, originariamente, si presenta come *homo faber*, e la tecnologia, fino dai tempi della pietra scheggiata, è come una protesi, che l'uomo congiunge ai propri arti per trasformare il mondo e dominarlo. Ma non tutto è così semplice, non tutto così roseo, non procedendo la storia umana, senza fratture né regressioni, verso le "magnifiche sorti e progressive" decantate dai laudatori del presente. E in più l'uomo non è solo *homo faber*, aggiungerei, ma detiene diversi altri predicati che lo rendono un animale complesso, a volte inquieto e insoddisfatto di essere livellato e ridotto alla mera "fabbrilità".

Già l'innesto fra dimensione tecnologica e dimensione umana è in cammino, tanto che definirei "tecnoumano" il nuovo *sistema*, ovvero il nuovo assieme che si profila: ma il problema è tutto sulla possibilità che

¹ Come nota la grande poetessa Emily Dickinson, la nostra mente «è più grande del cielo», nel senso che contiene, o può contenere, tutte le idee e perfino l'idea di infinito, affermazione adoperata come insegna preposta ai loro lavori da G. MAIRA, *Il cervello è più grande del cielo*, Solferino, Milano 2019 e G.M. EDELMAN, *Più grande del cielo*, Einaudi, Torino 2004.

l'uno di questi due elementi assuma il controllo della marcia in avanti; se, presa da «vergogna prometeica» di fronte alla sua debolezza di essere transitoria, l'umanità più o meno consapevolmente abdica, essa perderà il controllo su ciò che lei stessa ha programmato, divenendone una specie di schiava, e i prodotti accumulati le conferiranno in un primo tempo aiuto, rendendola però progressivamente più debole; se invece, con progressivi assalti di lucidità, l'uomo riprenderà il controllo del veloce cammino del mondo, allora riuscirà a imprimere la sua anima alla tecnologia, conferendo ad essa non solo un "supplemento" proveniente dalla sua umanità, ma proprio un'animazione integrale, quell'ispirazione cioè necessaria per creare una civiltà e per conferirle un sigillo ben definito².

In altri termini, o riusciamo a star di fronte all'innovazione tecnologica con creatività, o corriamo il rischio, davvero mortale, di soccombere ad essa e di esserne non coinvolti, ma travolti. Allora, in breve, la sfida mi sembra porsi in questi termini: tra i due atteggiamenti della demonizzazione e dell'idolatria, tra la via di chi vede solo mostri (anche perché non ha esperienza delle nuove tecnologie e le teme come si teme l'ignoto) e chi, altrettanto ingenuamente, fa di tali tecnologie un assoluto, sognando di "indiarsi" in maniera impossibile o di divenire immortale, la filosofia suggerisce una terza via: quella di "comprendere" la tecnologia, senza dimenticarne il significato strumentale, e di partecipare ai movimenti di umanizzazione. Occorre consapevolezza storica, non consegnando il passato e la memoria a tali strumenti tecnologici: sembra di sgravarci, invece ci svuotiamo.

I filosofi, con sobrietà e prudenza, dovrebbero far comprendere la necessità di un'etica per la civiltà tecnologica. Le ultime riflessioni di Remo Bodei sono basate sull'idea del *limite*, e non c'è idea più feconda, e più avversa alla nostra superbia e al nostro egocentrismo, che poi,

² Per la nozione di «vergogna prometeica» cf. G. ANDERS, *L'uomo è antiquato*, Bollati Boringhieri, Torino 2007; la locuzione «supplemento d'anima» è presente in un celebre scritto, amato da Adriano Olivetti: H. BERGSON, *Le due Fonti della Morale e della Religione* [1932], Edizioni di Comunità, Milano 1950.

a guardar bene, provengono dalla stessa radice: un esaltarsi, un voler slanciarsi al di sopra di sé, senza alcun limite. Avversario di tale prosopopea, il filosofo non deve limitarsi a scuotere la testa, e a lasciare che il mondo corra dove vuole, o dove è spinto dalla necessità, indicando invece la prospettiva di un vero e proprio *umanesimo del limite*, che faccia riannodare, fra l'altro, il *legame nuziale*, perduto e rimpianto, fra l'uomo e la "sua" Terra. Davvero, la Terra, valorizzata ma anche devastata, oggi «splende di trionfale sventura»³.

³ M. HORKHEIMER - T. W. [ADORNO, *Dialettica dell'illuminismo* 1944], Einaudi, Torino 2010, 11. Sul valore del "limite" cf. R. BODEI, *Limite*, il Mulino, Bologna 2016.

NOTE CONCLUSIVE

Gregorio Piaia

Al termine di un incontro così intenso e stimolante ringrazio vivamente i relatori di averci offerto una mole notevole di informazioni e idee per meglio accostarsi a questa realtà così proteiforme che va sotto il nome di Intelligenza Artificiale. Tentando di abbozzare una sintesi, va anzitutto detto che di fronte a tale fenomeno occorre evitare sia l'entusiasmo acritico di chi vede nell'IA la soluzione di problemi che sino a ieri apparivano insolubili (una sorta di panacea universale), in vista di una società "post-umana", sia la diffidenza o l'aperta avversione di chi dà risalto soprattutto agli effetti negativi. Non è certo una novità il fatto che tutte le scoperte e invenzioni che l'uomo ha compiuto si siano prestate e continuino a prestarsi a usi benefici e positivi ma anche decisamente negativi, e questo sin da quando un nostro remoto antenato capì che imbracciando un ramo di un certo spessore raccolto da terra poteva procurarsi frutti irraggiungibili con le sole mani e difendersi dalle aggressioni, ma poteva a sua volta aggredire non solo gli animali con cui cibarsi ma anche i suoi simili.

Non v'è dubbio che l'IA possa essere adoperata in maniera negativa, ma proprio per questo è necessaria una sua *regolamentazione* che risponda a istanze etiche (in primo luogo la salvaguardia della nostra libertà di pensiero e di azione) e non solo alle istanze del mercato o del potere politico. Si tratta di una sfida i cui esiti sono tutt'altro che scontati ma che va affrontata, anzitutto con una più adeguata conoscenza e valutazione della "cosa". Occorre infatti dissipare la convinzione che l'IA sia in assoluto una intelligenza più elevata di quella umana. In realtà la cosiddetta IA (il termine, in sé ambiguo, si è ormai imposto e non è facile mutarlo nell'uso corrente) non è altro che una "macchina" assai perfezionata, che in quanto tale va programmata e non è quindi in grado di prendere decisioni se non in base al programma e alle informazioni di

cui dispone. Che differenza c'è, allora, fra l'IA e le macchine calcolatrici a partire dalla celebre "Pascalina", inventata dal filosofo e matematico Blaise Pascal nella prima metà del Seicento? Rispetto alla tradizionale calcolatrice l'IA presenta non solo una memoria assai più potente e una elevatissima velocità di calcolo, ma – ecco il punto – ha un potere di *manipolazione* della nostra visione del mondo e del nostro comportamento che va ben oltre l'ambito della macchina come semplice strumento o "attrezzo", qual è, ad es. il termostato che regola il riscaldamento delle nostre abitazioni oppure l'auto che guidiamo correntemente e ai cui segnali elettronici è opportuno obbedire se vogliamo evitare incidenti.

Nell'ampia gamma delle applicazioni dell'IA vanno cioè distinti due livelli: il primo comprende tutti quei sistemi automatici che sono in grado di agire in maniera autonoma e "intelligente" per risolvere problemi di vario tipo; il secondo livello riguarda i sistemi che riproducono comportamenti umani e si affiancano al nostro modo di ragionare e quindi di decidere, e qui gli effetti comportamentali assumono un ruolo di primissimo piano. Un conto è infatti che tali sistemi siano di aiuto alla *persona* intesa come soggetto attivo e responsabile, un altro conto è che si sostituiscano al soggetto umano in maniera subdola con una rete di informazioni abilmente orientate oppure di *fake news*, come già da tempo sperimentiamo nel composito mondo del *web*. Il tutto a favore non di una crescita della capacità individuale di ragionare, giudicare e scegliere, ma per fare il gioco di un mercato ormai globalizzato o di gruppi di potere che dietro la veste esteriore poco o nulla hanno di democratico. V'è di più: quale sarà la sorte della stessa democrazia in una società in cui un sistema virtuale sottilmente oppressivo e impositivo finisca col prevalere sulle relazioni interpersonali, soffocando la capacità critica? Ritorna qui l'idea di una sfida che ci coinvolge tutti e alla quale non possiamo sottrarci, pena la rinuncia al nostro "essere persona", ovvero al nostro rapportarci agli *altri* (e all'*Altro*): un rapporto certamente complesso, mutevole e problematico ma vitale ed essenziale, in quanto ci costituisce come "umani", e al quale non dovrebbe subentrare il rapporto pseudo-personale con un *Super-altro* di natura elettronica.

Profili degli autori

ANDREA GALLUZZI è ingegnere elettronico-informatico, laureato in Filosofia all'Università di Perugia e in Ontologia trinitaria all'Istituto universitario Sophia di Loppiano (Figline e Incisa Valdarno), ove attualmente collabora per attività didattiche e di ricerca sul rapporto umano/tecnoscienze. Ha ricoperto vari ruoli nei settori *Information and Communication Technologies* e sviluppo software.

GIUSEPPE GOISIS, già professore ordinario di Filosofia politica e Politica ed etica all'Università Ca' Foscari di Venezia, studioso assai impegnato nell'associazionismo culturale (in particolare nella Fondazione "Luigi Stefanini"), ci ha lasciato nella notte fra il 4 e 5 aprile 2023. Pubblichiamo qui una sua riflessione che si lega strettamente al tema dell'IA e che è stata letta dalla consorte Monica nel corso del Convegno trevigiano.

MARKUS KRIENKE è 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 è 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) nonché avvocato abilitato al patrocinio avanti alla Corte di Cassazione. È direttore scientifico della rivista *Journal of Ethics and Legal Technologies* e fondatore di *CollectIUS*, piattaforma digitale di ricerca, discussione e archiviazione di casi di Tecnodiritto.

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

VITTORIO POSSENTI, già ordinario di Filosofia politica presso l'Università Ca' Foscari di Venezia, è stato membro del CNB, della Pontificia Accademia delle Scienze Sociali e di quella di San Tommaso d'Aquino. È presidente di "Persona al Centro. Associazione per la Filosofia della Persona". È autore di trentacinque volumi concernenti la metafisica e il nichilismo, la filosofia della persona, il pensiero politico, il nesso tra religione e civiltà.

RICCARDO POZZO è professore ordinario di Storia della filosofia all'Università di Roma Tor Vergata (2019), già direttore dell'ILIESI-CNR (2009-2012) e del DSU-CNR (2012-2017). Autore di: *Hegel: Introductio in Philosophiam* (La Nuova Italia 1989); *Kant und das Problem einer Einleitung in die Logik* (Lang 1989); *Georg Friedrich Meiers Vernunftlehre* (Frommann-Holzboog 2000);

Adversus Ramistas (Schwabe 2012); *History of Philosophy and the Reflective Society* (de Gruyter 2021).

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

LEOPOLDO SANDONÀ è direttore, oltre che docente stabile ordinario, presso l'ISSR di Vicenza e coordina il progetto "Etica e Medicina" della Fondazione Lanza. Autore di pubblicazioni in campo nazionale e internazionale, la sua ricerca verte sul pensare dialogico contemporaneo e sulle sue implicazioni, specie in ambito bioetico.

LUCIA STEFANUTTI, nata a Treviso, laureata a Ca' Foscari di Venezia in Storia della filosofia contemporanea, ha insegnato nei licei. Attualmente è presidente della Fondazione Stefanini e membro del Consiglio direttivo dell'associazione filosofica "Persona al Centro".

LUIGI VERO TARCA è professore onorario di Filosofia teoretica presso l'Università Ca' Foscari di Venezia, dove è stato anche direttore (poi emerito) del CESTUDIR (Centro Studi sui Diritti Umani). Oltre alle varie attività (ricerca, convegni, pubblicazioni) in Italia e all'estero, promuove da anni pratiche filosofiche orientate alla filosofia come stile di vita.

PUBBLICAZIONI DIGITALI – Triveneto Theology Press

FACOLTÀ TEOLOGICA IN RETE

SEDE A PADOVA

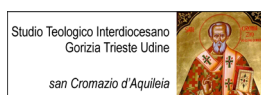


Ciclo Istituzionale
Ciclo di Specializzazione (Teologia pastorale e Teologia Spirituale)
Ciclo di Dottorato

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 (a cura), *Sulla fragilità dell'esistenza. Dialoghi*, 2023, pagine 84, isbn 979-12-81328-01-3.
- 2 - Antonio Bertazzo (a cura), *Quo vadis? Cammino, paradigma per Dio e per l'uomo*, 2023, pagine 68, isbn 979-12-81328-00-6.
- 3 - FRANCESCO PESCE (a cura), *Azione, esperienza, fede. Una prospettiva di teologia pratica*, 2023, pagine 193, isbn 979-12-81328-03-7.
- 4 - LIVIO TONELLO (a cura), *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*, pubblicazione 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, pubblicazione nel centenario della nascita di Luigi Sartori, 2024, pagine 66, isbn 979-12-81328-04-4.

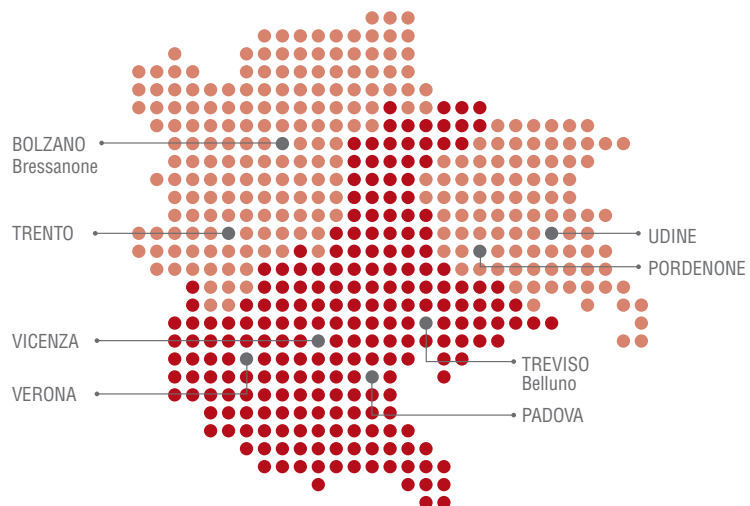
COLLANA EDUCATION

- 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.

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.

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

