SOPHIA

Ricerche su i fondamenti
e la correlazione dei saperi

Semestrale

Anno XVII n. 1
2025

SOPHIA *Rivista internazionale*
Istituto Universitario Sophia
Via San Vito 28, loc. Loppiano
50063 Figline e Incisa Valdarno (FI) – Italia
Anno XVII – 2025/1 (Gennaio-Giugno 2025)
ISSN 2036-5047

Direttore scientifico: Marco Martino

Comitato di redazione: Giuseppe Argiolas, Antonio Maria Baggio, Luigino Bruni, Bernhard Callebaut, Piero Coda, Valentina Gaudiano, Benedetto Gui, Marco Martino, Declan O'Byrne, Paul O'Hara, Giovanna Maria Porrino, Judith Povilus, Sergio Rondinara, Daniela Ropelato, Gérard Rossé, Marco Sanchioni.

Comitato scientifico internazionale: Angela Ales Bello (*Pontificia Università Lateranense, Roma*), Kurt Appel (*Università di Vienna, Austria*), Alessandra Beccarisi (*Università del Salento*), Maria Clara Lucchetti Bingemer (*Pontificia Università Cattolica di Rio de Janeiro, Brasile*), Vincenzo Buonomo (*Pontificia Università Lateranense, Roma*), Massimo Cacciari (*Università Vita-Salute San Raffaele di Milano*), Filipe Campello (*Universidade Federal de Pernambuco, Brasile*), Giuseppe Cantillo † (*Università di Napoli "Federico II"*), Peter Casarella (*Duke University School of Divinity, U.S.A.*), Bernhard Casper † (*Università Freiburg im Breisgau, Germania*), Claudio Ciancio (*Università del Piemonte Orientale*), Giuseppe D'Anna (*Università Cattolica del Sacro Cuore*), Mario De Caro (*Università degli studi Roma Tre, Roma*), Massimo Donà (*Università Vita-Salute San Raffaele, Milano*), Adriano Fabris (*Università di Pisa*), Emmanuel Falque (*Institut Catholique de Paris, Francia*), Riccardo Ferri (*Pontificia Università Lateranense, Roma*), Lorenzo Fossati (*Università Cattolica del Sacro Cuore*), Emmanuel Gabellieri (*Université Catholique de Lyon, Francia*), Gianluca Garelli (*Università di Firenze*), Giulio Giorello † (*Università degli Studi di Milano*), André Habisch (*Catholic University Eichstaett-Ingolstadt, Germania*), Vittorio Hösle (*University of Notre Dame, U.S.A.*), Philippe Hu (*Fu Jen Catholic University, Taiwan*), Luca Illetterati (*Università degli Studi di Padova*), Marco Ivaldo (*Università degli studi "Federico II" di Napoli*), Mario Longo (*Università degli Studi di Verona*), Daniel López (*Universidad Católica de Córdoba, Argentina*), Giancarlo Magnano San Lio (*Università di Catania*), Carmelo Meazza (*Università degli Studi di Sassari*), Mauro Mantovani (*Università Pontificia Salesiana, Roma*), Massimo Marassi (*Università Cattolica del Sacro Cuore, Milano*), Massimiliano Marianelli (*Università degli Studi di Perugia*), Giulio Maspero (*Pontificia Università Santa Croce di Roma*), Edoardo Massimilla (*Università degli studi "Federico II" di Napoli*), Letterio Mauro (*Università degli Studi di Genova*), Eugenio Mazzarella (*Università degli studi "Federico II" di Napoli*), John Milbank (*University of Nottingham, Inghilterra*), Donald W. Mitchell (*Purdue University, U.S.A.*), Juan Carlos Scannone † (*Facoltà di Filosofia e Teologia di San Miguel, Buenos Aires, Argentina*), Pierangelo Sequeri (*Facoltà Teologica dell'Italia Settentrionale di Milano*), Francesco Tomatis (*Università degli Studi di Salerno*), Roberto Tomichá (*Universidad Católica Boliviana de Cochabamba, Bolivia*), Giovanni Ventimiglia (*Università di Lucerna, Svizzera*), Vincenzo Vitiello (*Università Vita-Salute San Raffaele, Milano*), Lubomir Žak (*Università Palacký di Olomouc, Repubblica Ceca*), Stefano Zamagni (*Università degli Studi di Bologna*), Gonzalo Zarazaga (*Universidad Católica de Córdoba, Argentina*).

Segreteria scientifica: Raul Buffo (coordinatore), Tommaso Bertolasi, Alessandro Borghesi, Francesco Patrone
e.mail: rivista.sophia@sophiauniversity.org / raul.buffo@sophiauniversity.org

Rivista Semestrale Open Access
journal.sophiauniversity.org

Direttore responsabile: Michele Zanzucchi

Registrazione al Tribunale di Roma n° 405/2008
Iscrizione al R.O.C. n. 5849 del 10/12/2001

Gli scritti proposti per la pubblicazione in questa rivista sono peer reviewed

SOPHIA

* *Disegno della testata e progetto grafico di Ezio Bartocci*

SOPHIA

Ricerche su i fondamenti
e la correlazione dei saperi

SOPHIA *2025* INDICE

EDITORIALE

9 Che cos'è "pensare", oggi?
Una riflessione interdisciplinare su intelligenza artificiale e condizione umana
Marco Martino

SAGGI

13 L'intelligenza artificiale generativa e il gioco dell'imitazione
Vincenzo Fano

27 Agli albori dell'AI. Riflessioni sul test di Turing
Francesco F. Calemi

41 Virtù Intellettuali e Intelligenza Artificiale: il *Machine Learning* può fornire una Comprensione Scientifica del mondo?
Vera Matarese

55 L'innovazione tecnologica dell'IA nell'età della tecnica
Sergio Rondinara

69 Il ruolo della transdisciplinarità nell'era dell'Intelligenza Artificiale (Generativa)
Marco Sanchioni

95 Vulnerabilità creativa e intelligenza artificiale
Antonio Bergamo

LABORATORIO

109 Intelligenze Artificiali e la disuguaglianza di genere. Criticità e prospettive
Valentina Gaudiano

123 Artificial and Human Intelligence: Algorithmic Bias and the Myth of Neutrality
Seydel Ducosquel

133 AI and Women: Between Automation and Meaningful Work
Aiza Asi

RICERCHE

147 Uno sguardo musulmano sulla tecnologia: la prospettiva di Mohammed Aziz Lahbabi
Francesco Patrone

FORUM

159 Sulla luce della Rivelazione. A proposito di *L'*eclair *di una Rivelazione* di Carmelo Meazza
Alberto Cosseddu

165 Aprire nuovi orizzonti di vita con Bergson. A partire da *Avec Bergson* di Frédéric Worms
Francesco Patrone

EDITORIALE

Che cos'è "pensare", oggi? Una riflessione interdisciplinare su intelligenza artificiale e condizione umana

Nel 1950 Alan Turing scriveva: *Le macchine possono pensare?* Oggi, a settant'anni di distanza, ci troviamo a riformulare la domanda: Che cosa significa pensare, nell'era delle macchine? In un'epoca segnata da transizioni profonde e accelerate, la questione dell'intelligenza artificiale si impone non solo come tema tecnico-scientifico, ma come nodo/snodo filosofico, antropologico, economico-politico e spirituale. Un invito dunque a riconsiderare le categorie fondamentali della soggettività nell'epoca delle macchine che apprendono, rispondono e simulano intenzionalità. "Pensare" non è solo calcolare, né semplicemente produrre risposte coerenti. È capacità di giudizio, interiorità riflessiva, apertura all'alterità e all'imprevisto. È l'atto di porre domande prima ancora che di formulare risposte, il gesto con cui il soggetto si espone al reale, anche quando questo disattende le sue previsioni. Se "pensare" è solo trattare informazioni, allora le macchine già pensano; ma se è anche esperienza del limite, capacità di sospensione e immaginazione di mondi possibili, allora la posta in gioco è più radicale.

I sistemi di intelligenza artificiale generativa, ormai capaci di dialogare, imitare, suggerire, produrre e talvolta persino sorprendere, ci pongono davanti a un bivio: adattarci passivamente all'automazione dei linguaggi e dei gesti, o intraprendere un processo più esigente e creativo di riflessione, discernimento e rigenerazione culturale. Questo numero della rivista Sophia nasce dall'urgenza di interrogare l'intelligenza artificiale non solo come fenomeno tecnologico, ma come opportunità per la comprensione dell'umano, alla luce delle trasformazioni epistemiche e culturali in atto.

Nel solco di questo interrogativo fondamentale, si inserisce idealmente anche il dialogo avviato fin dal numero 0 della rivista Sophia, dove Giuseppe Maria Zan-

ghí, con l'articolo "*Che cos'è pensare*?"[1], offriva una riflessione profonda e anticipatrice su ciò che distingue l'atto umano del pensare da ogni sua possibile imitazione. In quel testo, il pensare veniva delineato come gesto originario, sorgivo, capace di generare relazione e significato, oltre ogni mera funzionalità. Un invito, ancora oggi attuale, a custodire la densità dell'umano proprio là dove sembra farsi più evanescente. Pensare, oggi, significa forse anche resistere all'automatismo del pensiero stesso: è l'arte di tenere aperta una domanda, là dove tutto sembra già risolto[2].

In ambito accademico, si moltiplicano gli approcci che mettono in dialogo neuroscienze, epistemologia, filosofia della tecnica e teoria della comunicazione, nel tentativo di interrogare in modo più articolato le implicazioni ontologiche, cognitive e sociali dell'intelligenza artificiale. Questo intreccio transdisciplinare non nasce da un semplice interesse per l'innovazione tecnologica, ma da un'urgenza teorica: comprendere in che modo l'IA stia ridefinendo le nostre categorie fondamentali di soggetto, azione e conoscenza.

Da un lato, si assiste a un rinnovato interesse per le categorie classiche dell'antropologia filosofica – libertà, responsabilità, finitezza – rilette alla luce delle sfide poste dagli algoritmi e dalla crescente automazione delle scelte. Dall'altro, emergono prospettive che mirano a una ridefinizione del concetto stesso di "intelligenza", orientandolo verso modelli relazionali, situati e contestuali, capaci di valorizzare la pluralità dei modi di conoscere e agire nel mondo, e di integrare dimensioni spesso escluse dalla razionalità computazionale, come l'intuizione, l'empatia, il corpo e il vissuto. In questo scenario, la sfida non è soltanto teorica: essa interpella direttamente le pratiche educative e formative, chiamate a coltivare menti capaci non solo di utilizzare l'IA in modo efficace, ma anche di comprenderne criti-

1 - Cfr. G.M. Zanghí, *Che cos'è pensare?*, in «Sophia. Ricerche sui fondamenti e la correlazione dei saperi», (2008/0), pp. 25-34; cfr. anche il numero monografico 2022/2 della rivista «Sophia», dedicato al tema del pensare.

2 - A mo' di esempio – e come segnale di una domanda filosofica originaria tuttora aperta – vale forse la pena soffermarsi su ciò che, anche per le più avanzate forme di IA, continua a rimanere opaco: l'*istante*, l'irruzione dell'"*improvviso*". Platone, nel *Parmenide*, chiama ἐξαίφνης quel passaggio liminale, non collocabile nel tempo o nello spazio, in cui il mutamento accade – tra il "non più" e il "non ancora". È lì che qualcosa *si dà*, senza essere previsto né prodotto. Nell'istante si accende, come scrive Platone nella *Lettera VII*, la conoscenza filosofica, come una scintilla che nasce "d'improvviso" nell'anima e si alimenta da sé. Nessuna macchina può *attendere* quell'imprevisto, né *riconoscere* nel lampo una rivelazione. L'intelligenza artificiale può imitare il linguaggio, ma non conosce il balbettio del "non così" che accompagna ogni vera trasformazione interiore. In un'epoca segnata dall'efficienza, forse il compito più radicale è proprio questo: custodire lo spazio per ciò che *non si lascia automatizzare*, per l'evento che *sfugge al codice* e apre l'umano alla meraviglia.

camente i limiti, gli impatti sistemici e soprattutto le implicazioni antropologiche, culturali e politiche. Questa complessità crescente non va semplicemente celebrata come un superamento delle dicotomie tradizionali tra umano e artificiale, tra naturale e tecnologico; al contrario, richiede un lavoro di discernimento critico, capace di abitare le tensioni senza dissolverle.

È proprio nel confronto con tali polarità che si apre, forse, una possibilità: non una fusione indifferenziata, ma un riconoscimento dei confini, delle ambivalenze e delle domande che esse generano. In questo senso, la riflessione accademica è chiamata a custodire lo spessore della distinzione – tra intenzionalità e automazione, tra comprensione e simulazione – senza rinunciare alla ricerca di nuove sintesi, che non cancellino ma rilancino la profondità e la complessità dell'umano.

Il fascicolo si apre con un importante contributo di Vincenzo Fano, che rilegge il test di Turing alla luce delle più recenti applicazioni dell'IA generativa, evidenziando come la rinuncia alla definizione essenziale del pensiero abbia reso possibile una nuova epistemologia dell'intelligenza. La questione, spostata dal "pensare" al "comportarsi come se si pensasse", rivela l'ambiguità di una macchina che può imitare l'umano senza comprenderlo, e tuttavia sollecita la filosofia della mente a fare i conti con una soglia superata: quella dell'indistinguibilità operativa.

Su questa soglia si colloca anche il saggio di Francesco Federico Calemi, che ritorna all'articolo fondativo di Turing ricostruendone le premesse e il contesto, ma anche confrontandolo con l'eredità di Leibniz e con i dilemmi classici del funzionalismo. Il "gioco dell'imitazione" è qui interpretato come un dispositivo che mette alla prova non solo la macchina, ma soprattutto il nostro modo di attribuire significato, intenzionalità e interiorità.

Dalla filosofia alla psicologia, dalla scienza dei dati alla cultura, l'itinerario prosegue con il contributo di Vera Matarese, che si interroga sulla possibilità che il *machine learning* possa costituire una forma di comprensione scientifica del mondo, affrontando il delicato rapporto tra virtù intellettuali e automazione cognitiva. A sua volta, Sergio Rondinara propone una riflessione sull'innovazione tecnica dell'IA, inscrivendola nel quadro più ampio dell'età della tecnica, là dove la ragione strumentale rischia di eclissare la dimensione assiologica dell'agire umano.

Un elemento ricorrente nei vari contributi è la necessità di un approccio transdisciplinare, tema affrontato esplicitamente da Marco Sanchioni, che sottolinea come solo una convergenza di saperi – scientifici, filosofici, educativi – possa generare una vera intelligenza al servizio dell'umano. In questa linea si inserisce anche la prospettiva offerta da Antonio Bergamo, che introduce la nozione di "vulnerabilità creativa" come chiave per comprendere il rapporto tra l'umano e l'artificiale, suggerendo che la capacità di fallire e rigenerarsi resti prerogativa essenziale dell'uomo. In un contesto dominato dall'efficienza e dalla predittività, la vulnerabi-

lità – lungi dall'essere un difetto – diventa cifra dell'umano. È nella possibilità di fallire, apprendere e trasformarsi che si custodisce la dignità dell'esperienza. La macchina, per quanto sofisticata, non conosce la ferita, né il riscatto.

Il fascicolo ospita inoltre una sezione "laboratorio" dedicata all'intreccio tra IA e questioni di genere, con il contributo di Valentina Gaudiano che esplora le implicazioni etiche e sociali della disuguaglianza algoritmica, mettendo in evidenza come gli stessi sistemi di intelligenza, se non progettati consapevolmente, possano perpetuare stereotipi e ingiustizie. Su questo sfondo si collocano anche le riflessioni di Seydel Ducosquel sul mito della neutralità algoritmica e quelle di Aiza Asi, che si concentrano sul lavoro femminile nel tempo dell'automazione, mostrando come il concetto stesso di "lavoro significativo" debba essere ripensato alla luce delle trasformazioni tecnologiche.

La sezione Ricerche amplia ulteriormente l'orizzonte, accogliendo il contributo di Francesco Patrone, che propone uno sguardo musulmano sulla tecnologia attraverso la figura del filosofo marocchino Mohammed Aziz Lahbabi, in un dialogo che intreccia spiritualità, tecnica e umanesimo.

Infine, il fascicolo si chiude con due recensioni che completano il quadro aperto dal tema principale: una rilettura del volume *L'*eclair *di una Rivelazione* di Carmelo Meazza a cura di Alberto Cosseddu, e una riflessione di Francesco Patrone sul testo *Avec Bergson* di Frédéric Worms. Entrambe le recensioni, sebbene collocate nella sezione Forum, rappresentano un ulteriore invito ad aprire nuovi orizzonti di pensiero, in continuità con la vocazione della rivista.

In un mondo che si interroga su che cosa significhi "pensare", "decidere" e "creare" in presenza di macchine capaci di calcolo e apprendimento, *Sophia* ribadisce la necessità di un sapere critico, radicato, capace di visione. La "rinascita dell'umano", in questo senso, non è nostalgia ma compito: implica discernimento, dialogo tra i saperi e un'educazione al limite come luogo generativo. È solo nella misura in cui sapremo abitare questi confini – tra scienza e coscienza, calcolo e cura, efficienza e empatia – che l'intelligenza artificiale potrebbe diventare risorsa, e non rischio.

Non è la macchina, forse, a dover diventare più umana, ma l'umano a riscoprire ciò che lo rende tale.

MARCO MARTINO
Professore associato di Storia della filosofia presso il Dipartimento di Filosofia, Scienze sociali, umane e della formazione dell'Università degli Studi di Perugia
marco.martino@unipg.it

SAGGI

L'intelligenza artificiale generativa e il gioco dell'imitazione

di Vincenzo Fano

This paper discusses the recent experiments by Jones and Bergen that showed ChatGPT-4 passed the Turing test. To this end, the game proposed by Turing is placed in the proper philosophical dimension by showing how it tries to answer a well-formulated and circumscribed question. The details of the experiment are then investigated. Finally, some important consequences are drawn regarding the problem of explainability of neural networks.

Keywords: Turing Test; ChatGPT-4; Explainability of Neural Networks; Imitation Game; Philosophy of Psychology

Introduzione

Il tentativo di sviluppare l'intelligenza artificiale in modo non logico ha origini coeve rispetto al progetto mainstream basato su modelli logici[1]. "Non logico" nel senso duplice di basarsi sull'addestramento e sull'apprendimento più che su un software o un hardware già strutturati e prendere le mosse da una configurazione semplice e simmetrica, che poi si sviluppa mediante l'addestramento. Questo approccio, che proviene più dalla psicologia che dall'informatica, venne aspramente criticato in un celebre libro del 1968[2]. La critica, pur corretta, non teneva in considerazione il fatto che il percettrone poteva essere sviluppato con gli strati intermedi, come venne poi proposto sempre dagli psicologi, negli anni Ottanta[3]. I limiti di quest'approccio, ispirato in parte dal funzionamento del nostro cervello, erano invece altri due: in primo luogo, la scarsità di dati etichettati e in secondo luogo, la limitata potenza dei calcolatori. Solo nel 2012, AlexNet ottenne alla celebre competizione ImageNet un risultato strabiliante – 95% di immagini classificate correttamente. AlexNet era una rete neurale convolutiva, basata su un'idea di Ian LeCun, realizzata da Geoffrey Hinton, Alex Krizhevsky e Ilya Sutskever: tutti nomi che hanno fatto la storia della rivoluzione tecnologica a cui stiamo assistendo.

Da quel momento in poi le reti neurali addestrate sono entrate nella nostra vita quotidiana in modo sempre più capillare. L'idea non logica degli psicologi ha di fatto vinto rispetto a quella logica proposta dalla maggior parte degli studiosi di intelligenza artificiale. Vedremo però che in un certo senso la logica si prenderà la rivincita.

Di solito si fa risalire la fondazione del progetto dell'intelligenza artificiale al famoso lungo congresso del 1956 a Dartmouth, durante il quale venne coniato il termine a cui parteciparono studiosi del calibro di Claude Shannon, Marvin Minsky e Herbert Simon. D'altra parte, nessuno può negare che questo nuovo programma di ricerca abbia avuto fra le pagine del celebre articolo di Alan Turing comparso sulla rivista filosofica *Mind* nel 1950 una spinta decisiva[4]. Questo straordinario saggio, fra le altre cose, proponeva quello che venne poi chiamato "test di Turing", cioè un gioco ideato come cartina di tornasole per stabilire se le macchine possano o meno pensare.

---

1 - W.S. McCulloch - W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, «Bulletin of Mathematical Biophysics», 5 (1943), pp. 115-133 e F. Rosenblatt, *The perceptron: a probabilistic model for information storage and organization in the brain*, «Psychological review», 65 (1958/6), p. 386.

2 - M. Minsky - S.A. Papert, *Perceptrons*, MIT Press, Cambridge MA 1968.

3 - D.E. Rumelhart - J.L. McLelland (eds.), *Parallel, Distributed Processing,* MIT Press, Cambridge MA 1986, Vol. I e II.

4 - A. Turing, *Computing machinery and intelligence*, in «Mind», 59 (1950), pp. 433-460.

Come spiega Melanie Mitchell nel suo importante libro del 2019[5], il test di Turing non è preso troppo sul serio dagli informatici, in quanto si presta a trovate pubblicitarie di dubbio valore tecnologico. Inoltre, l'intelligenza è tante altre cose, non solo superare il test di Turing. Infine, anche chatbot estremamente poco sofisticati riescono a simulare conversazioni umane.

Con l'avvento delle nuove reti neurali, come la serie di ChatGPT, il test di Turing è tornato a far parlare di sé. Questa tecnologia è molto complessa e simula in modo sorprendente la capacità di conversare umana. Nello Cristianini, nel suo bel libro, *Machina Sapiens*[6], racconta che gli stessi ingegneri e programmatori che la hanno realizzata sono rimasti sorpresi dalle sue abilità di dialogo. Resta però il fatto che spesso si è ancora scettici sulla capacità di ChatGPT di superare il test di Turing. Ad esempio, Tommaso Poggio, nel suo ottimo volume, *Cervelli, menti, algoritmi*, scritto con Marco Magrini, sostiene che «i *large language models* […] almeno per ora non sono in grado di superare una severa versione del test di Turing»[7]. "Per ora" nel 2023, ma già nel 2024 sono usciti due articoli che descrivono esperimenti molto seri, dai quali risulta, invece, che ChatGPT-4 ha superato ampiamente il test di Turing[8]. Di questo parleremo nel presente breve testo. Nel prossimo paragrafo delineeremo una metodologia generale in filosofia della psicologia che consente di collocare correttamente il significato filosofico del test di Turing. In quello successivo descriveremo e discuteremo il gioco dell'imitazione proposto da Turing nel 1950. Nella sezione 4 analizzeremo gli esperimenti da cui risulta che il test è stato superato. Alcune brevi considerazioni concludono l'articolo.

Il quadro generale

Se decidiamo di provare a comprendere qualcosa degli elettroni, dobbiamo studiare la meccanica quantistica, cioè la migliore teoria che possediamo riguardo a queste particelle. "Migliore" nel doppio senso che è molto ben confermata e altamente esplicativa. Se accettiamo almeno in parte la tesi del realismo scientifico, allora, pur consapevoli che la teoria quantistica è una creazione umana, siamo convinti che essa dica anche qualcosa su come sono effettivamente fatti gli elettroni. In che senso possiamo essere realisti scientifici moderati non lo discutiamo in que-

5 - M. Mitchell, *Intelligenza artificiale,* Einaudi, 2022 pp. 38-41. L'ed. originale è del 2019.

6 - N. Cristianini, *Machina Sapiens,* Il Mulino, Bologna 2024

7 - T. Poggio - M. Magrini, *Cervelli, menti, algoritmi*, Sperling & Kupfer, Milano 2023, p. 111.

8 - C.R. Jones - B. K. Bergen, *Does GPT-4 pass the Turing test?*, in «*arXiv*» 2023: *arXiv:2310.20216*; C.R. Jones - B.K. Bergen, *People cannot distinguish GPT-4 from a human in a Turing test, in* «arXiv» 2024: *arXiv:2405.08007.*

sto contesto[9]. Basta dire che la meccanica quantistica è almeno in parte vera. Dove il termine "vera" va inteso in senso corrispondentista.

Stiamo indugiando sull'epistemologia di una scienza naturale altamente confermata, al fine di chiarire alcuni punti che torneranno utili rispetto al nostro problema, che in fondo è di filosofia della psicologia.

All'interno di un paradigma genericamente aristotelico di scienza, per comprendere gli elettroni dovremmo stabilire quali sono le loro caratteristiche essenziali – quella che potremmo chiamare "sostanza seconda" dell'elettrone – e allo stesso tempo fornire una definizione che ne colga la forma, cioè la "sostanza prima". La prima istanza fa ancora parte della scienza moderna. Un elettrone ha una certa massa e una certa carica, che lo contraddistinguono, ma queste sue peculiarità di fatto non dicono quasi nulla del reale suo comportamento; la seconda è stata completamente abbandonata. Infatti, conoscere gli elettroni significa oggi stabilire le equazioni che governano il loro comportamento nelle diverse situazioni, come ad esempio, quando l'elettrone è libero o fa parte di un atomo di idrogeno. Avere a disposizione queste equazioni, tuttavia, non ha nulla a che fare con sapere l'essenza dell'elettrone; significa invece, conoscerne il comportamento e saperlo manipolare.

Questo esempio mostra un punto molto importante che caratterizza l'epistemologia della scienza moderna in generale: non proviamo più ad afferrare l'essenza della realtà, ma solo a comprenderne alcuni comportamenti e a fornirne delle spiegazioni funzionali.

Dobbiamo tenere presente questa scelta metodologica in parte rinunciataria, ma vincente, quando prendiamo in considerazione lo studio della mente umana.

La psicologia nasce come scienza sperimentale nella seconda metà dell'Ottocento. Di solito come data discriminante si fa riferimento all'istituzione a Lipsia del primo laboratorio di psicologia sperimentale fondato da Wilhelm Wundt nel 1879[10]. A grandi linee, Wundt sostiene che ci sia un piano dell'esperienza vissuta precedente al costituirsi dei concetti scientifici. La psicologia sarebbe quindi la scienza empirica e sperimentale di questo livello di realtà primario, basata sull'auto-osservazione. Questo progetto epistemologico ha dato qualche frutto, ma è rapidamente andato incontro a problemi insolubili. Già Comte aveva osservato che la psicologia intesa in questo senso è impossibile, dato che l'osservatore e l'osservato coincidono e quindi interferiscono[11]. Inoltre Wittgenstein mostra che di fatto una scienza

9 - Vedi però V. Fano, *A Regional Scientific Image of Time*, in «Argumenta», 8 (2/2023), pp. 407-425 [https://www.argumenta.org/article/a-regional-scientific-image-of-time-special-issue/].

10 - Vedi, ad esempio, E.G. Boring, *A history of experimental psychology*, Appleton-Century-Crofts, New York 1957[2].

11 - A. Comte, *Cours de philosophie positive*, Bachelier, Paris 1864[2], vol I, pp. 30ss.

dei propri vissuti è impossibile, poiché il linguaggio pubblico e condiviso li modifica e li costituisce[12]. L'introspezione dunque in generale si dimostra un metodo tutt'altro che affidabile per fare psicologia scientifica. L'idea di una scienza prima dei vissuti passa progressivamente alla filosofia. In varie forme la sosterranno Bergson, Dilthey, Mach, Husserl e Carnap[13], fra gli altri. D'altra parte, la psicologia, al fine di indirizzarsi su una base più scientifica, soprattutto negli Stati Uniti, si avvicina a posizioni comportamentiste, come quella prima di Watson e poi di Skinner. Questi ultimi abbandonano completamente il progetto di mettere a punto una scienza della soggettività, concentrandosi invece sugli stimoli e sui comportamenti, cioè sulla parte pubblica e condivisa del fenomeno psichico.

Negli anni Sessanta del Novecento il modello comportamentista si dimostrerà insufficiente. Tuttavia questo non darà origine a un ritorno della psicologia come scienza dell'esperienza soggettiva, bensì all'introduzione di modelli mentali non osservabili che intervengono a spiegare l'elaborazione umana degli stimoli e la produzione di comportamenti. Questi modelli possono essere suggeriti dai report dell'esperienza vissuta da parte dei soggetti sperimentali, ma possono essere anche, come si dice oggi, subpersonali, cioè di fatto inconsci. Anzi, per la maggior parte sono modelli di come lavorerebbe il nostro cervello, che avrebbero poco a che fare con quello che percepiamo soggettivamente, che, come abbiamo accennato, è sfuggente e non facilmente oggettivabile. Inoltre negli ultimi decenni abbiamo acquisito molti nuovi dati sul funzionamento neurofisiologico del nostro cervello, per cui è oggi anche possibile confrontare, almeno in parte, i modelli cognitivi con la nostra neurofisiologia. Resta il fatto che la psicologia scientifica non è più una scienza della nostra esperienza soggettiva[14].

Questa situazione generale ha molte eccezioni, su cui vale la pena soffermarsi brevemente. Ad esempio, Gerald Edelman, nel suo classico libro *La materia della mente*[15], prova a sostenere l'esigenza di una scienza "non galileiana", cioè che provi a tenere conto anche dell'esperienza soggettiva. Il progetto neuro-fenomenologico di Varela[16] si muove nella stessa direzione. Recentemente Dan Zahavi ha provato a riproporre una psicologia fenomenologica[17]. I gestaltisti, che ormai sono una spa-

12 - L. Wittgenstein, *Ricerche filosofiche*, Einaudi, Torino 2009. L'ed. originale è del 1953.

13 - Vedi ad esempio H. Spiegelberg, *The phenomenological movement*, Springer, Berlin 1971 e W. Stegmüller, *Die Hauptströmungen der Gegenwartsphilosophie*, Kröner, Stuttgart 1965.

14 - M.R.W. Dawson, *Mind, Body, World. Foundations of cognitive science*, AU Press, Edmonton AB 2013.

15 - G. Edelman, *La materia della mente*, Adelphi, Milano 1992.

16 - F.J. Varela, *Neurophenomenology: a methodological remedy for the hard problem,* in «Journal of Consciousness Studies», 4 (3/1996), pp. 330-349.

17 - S. Gallagher - D. Zahavi, *The Phenomenological Mind*, Routledge, London 2008.

ruta minoranza, non hanno mai abbandonato il progetto di tenere conto dell'esperienza soggettiva, soprattutto quando si tratta di psicologia della percezione; ecc.

Tutti questi tentativi di oggettivare l'esperienza soggettiva hanno ottenuto risultati parziali e controversi. Proviamo a indagare meglio le ragioni di questa difficoltà.

Al fine di comprendere questo problema, torniamo brevemente all'epistemologia della fisica. Fin dai tempi di Newton questa scienza ha introdotto per ragioni esplicative entità non direttamente misurabili, come le forze, i campi e le particelle. Questo significa che in psicologia non dobbiamo avere paura a fare altrettanto. E in effetti, il cognitivismo, come abbiamo visto brevemente prima, si è mosso proprio in questa direzione. Le entità fisiche non misurabili hanno due caratteristiche salienti: in primo luogo, sono descrivibili in maniera esatta, la maggior parte delle volte mediante un linguaggio matematico; in secondo luogo, sono connesse a quelle misurabili da precise leggi funzionali o indirettamente o direttamente. Qualcosa di simile capita per i modelli cognitivi, che sono spesso formulati in termini logici o computazionali e sono ben connessi agli stimoli e ai comportamenti che vorrebbero spiegare.

Quando invece passiamo all'esperienza soggettiva, tutto cambia. Abbiamo buone ragioni epistemologiche per ritenere che le altre persone abbiano un'esperienza soggettiva simile alla nostra. In effetti le altre persone sono molto simili a noi dal punto di vista biologico; noi abbiamo un'esperienza soggettiva, quindi viene da dire che anche loro la abbiano. È vero che questa è una strana inferenza a partire da un solo caso, tuttavia la nostra esperienza soggettiva ci accompagna per buona parte della nostra vita, è fortemente correlata a ciò che capita al nostro corpo e l'analogia fisico-biologica fra noi e gli altri è molto dettagliata. Possiamo dunque concludere che gli altri vivano un'esperienza soggettiva simile alla nostra. A questo punto siamo, dal punto di vista epistemologico, in una situazione che ha vantaggi e svantaggi rispetto a quella della fisica. In primo luogo, non possediamo un linguaggio esatto per descrivere la nostra esperienza soggettiva, come la matematica o la logica. Tuttavia, in un certo senso, anche se deformato dal linguaggio e dalle nostre idiosincrasie, abbiamo, almeno in prima approssimazione un accesso privilegiato alla realtà della nostra soggettività, che, invece, nel caso, ad esempio, di un elettrone è del tutto assente. Questo accesso parzialmente privilegiato è un'arma a doppio taglio: da un lato ci convince che non sia così difficile cogliere l'essenza dei fenomeni mentali, dall'altro, però, non abbiamo alcun modo per oggettivarli. Infatti, non solo ci manca un linguaggio adeguato, come la matematica o la logica, ma non siamo neanche in grado di dire in modo chiaro che cosa sarebbero dei contenuti soggettivi oggettivati. Ad esempio, che significato potrebbe avere "percepire la mente di un'altra persona"? Se stiamo percependo, stiamo percependo un nostro stato mentale, non quello di un altro. Inoltre, che senso avrebbe l'espressione "so quel che pensi", se di fatto non possediamo un linguaggio ogget-

tivo per descrivere i nostri pensieri? Insomma, al di là dell'ingannevole intuizione che avremmo un accesso privilegiato agli stati mentali, fare una scienza della soggettività sembra un compito oltremodo difficile.

Questo non significa che sia del tutto impossibile. Da un lato notiamo che per conoscere non esiste solo la scienza, ma anche altre attività tipicamente umane. Come hanno mostrato, ad esempio Nelson Goodman e Hans-Georg Gadamer[18], anche l'arte insegna. Certo sono conoscenze con un basso grado di esattezza e giustificazione, ma non sono del tutto assenti. In altre parole, raccontare storie sulla nostra soggettività è tutt'altro che inutile, anche se non può avere i crismi della scientificità. Dall'altro, in un certo senso la nostra esperienza soggettiva, anche se non in modo diretto e costitutivo, gioca un ruolo indiretto e regolativo nella psicologia scientifica. Faccio un esempio. Chiaramente non possiamo utilizzare una nostra esperienza soggettiva per giustificare un asserto psicologico; tuttavia il nostro vissuto ci guida nell'indagine dei comportamenti e nella costruzione dei modelli.

Questa prima parte sull'epistemologia della psicologia scientifica ci è servita per introdurre il test di Turing. Come vedremo nella prossima sezione, questo test è rinunciatario dal punto di vista epistemologico, ma è proprio ciò di cui abbiamo bisogno per fare buona scienza. In psicologia, esattamente come in fisica, dobbiamo seguire il principio galileiano di "non tentar l'essenza".

Il gioco dell'imitazione

Alan Turing pubblica sulla rivista filosofica *Mind* nel 1950 un articolo straordinario, che ancora oggi influenza il nostro modo di concepire i computer e il loro rapporto con la mente umana. Il titolo del saggio è *Macchine calcolatrici e intelligenza*. Di fatto però Turing si pone una domanda diversa: "possono le macchine pensare?" Non sappiamo se Turing fosse consapevole della distinzione kantiana fra "intelletto" e "pensiero". Il primo sovraintende l'attività argomentativa della mente umana, mentre il secondo ha un senso molto più ampio, che comprende anche l'immaginazione e le idee della ragione. Lo sviluppo del saggio e degli esempi che l'Autore propone portano a ritenere che Turing abbia in mente proprio la nozione ampia di pensiero. Tuttavia questa domanda così formulata è troppo ambigua. Turing propone quindi una questione molto diversa, che è indagabile e precisa. Mettiamo in due stanze chiuse una macchina e una persona. Dopo di che un'altra persona, chiamata l'interrogante, può comunicare con le due stanze solo mediante messaggi scritti a macchina. Dopo un po' di dialogo l'interrogante deve stabilire in

18 - N. Goodman, *I linguaggi dell'arte,* Il Saggiatore, Milano 1976 e H.-G. Gadamer, *Verità e metodo*, Bompiani, Milano 1983.

quale stanza ci sia la macchina e in quale la persona. Di fatto la persona prova ad aiutare l'interrogante nel suo compito, mentre la macchina non deve farsi scoprire.

Dunque, la domanda, "possono le macchine pensare?" viene sostituita da "può una macchina ingannare l'interrogante nel gioco dell'imitazione?". Prima di proseguire, sottolineiamo l'aspetto metodologico centrale della mossa di Turing: egli ha evitato di rispondere alla domanda troppo ambiziosa "che cosa è il pensiero?". Come abbiamo già detto, oggi non ci chiediamo "che cosa è un elettrone?", ma proviamo a spiegare e prevedere come esso si comporta. Analogamente, noi non siamo in grado di rispondere in modo univoco alla domanda "che cosa è il pensiero?"; però sappiamo come si comportano le persone che pensano. Nella visione di Turing, pensare è un'attività prettamente intellettuale che si esprime linguisticamente.

Oggi, la scienza cognitiva, dopo aver accettato questa idea per decenni, la sta mettendo in discussione, attribuendo un'importanza sempre maggiore al corpo e all'azione[19]. Probabilmente Turing era consapevole che la sua nozione prettamente linguistica di pensiero fosse limitata, ma taglia corto: questa è la domanda che vuole discutere, senza entrare nel merito se sia o meno quella giusta. Per Turing, dunque, il pensare coincide con i comportamenti linguistici di una persona. Notiamo anche che Turing si disinteressa completamente di come la macchina riesca a simulare i comportamenti linguistici umani. Essa può utilizzare hardware e software sostanzialmente diversi da quelli di cui si avvalgono le persone. Attualizzando il suo modello, il fatto che oggi le immense reti neurali dietro ai chatbot funzionino in modo sostanzialmente diverso dal cervello umano è irrilevante; l'importante è che i comportamenti linguistici vengano effettivamente simulati in modo adeguato.

È importante anche sottolineare che Turing non attribuisce alcun ruolo alla soggettività e alla coscienza. Abbiamo visto nella sezione precedente che non dobbiamo farci ingannare dal nostro accesso ai nostri stati mentali. Infatti, il mondo in prima persona, pur essendo molto importante, non è oggetto di scienza. Qualche pagina dopo, Turing sarà categorico: «Il solo modo di conoscere che un uomo pensa è essere quell'uomo»[20]. In altre parole, se volessimo coinvolgere la coscienza nella nostra discussione scientifica della nozione di pensiero saremmo costretti a una visione solipsistica. Questo non significa, prosegue Turing, che non ci siano misteri riguardo alla coscienza, ma solo che ci si può porre la domanda sul pensiero delle macchine anche senza aver risolto quei misteri. In altre parole, come spesso accade nella metodologia scientifica, la strategia *divide et impera* è vincente.

---

19 - Vedi ad esempio, M.R.W. Dawson, *Mind, Body, World: Foundations of Cognitive Science*, Athabasca University Press, Edmonton AB 2013, cap. 5.

20 - A.M. Turing, *Computing machinery and intelligence*, in «Mind» 59 (1950), pp. 433-460, qui p. 446. Trad. nostra.

Da questa presa di posizione di Turing segue che tutti quegli argomenti, come la "stanza cinese" di Searle o la "testa di legno" di Block[21], sono sostanzialmente irrilevanti. Per essere più precisi, la nozione di pensiero di cui Turing si vuole occupare non comprende la soggettività, che resta un concetto sostanzialmente estraneo al suo approccio parzialmente operazionale.

Turing prosegue spiegando che cosa intende con il termine "macchine" nella sua domanda. Egli si concentra sulle cosiddette "macchine digitali" universali. Turing non è esplicito al riguardo, ma sembra di capire che egli abbia in mente un computer che sia una macchina universale di Turing, cioè che sia in grado di simulare tutte le possibili computazioni nel senso di Turing.

Resta aperta la questione delle macchine non digitali, come ad esempio il regolo calcolatore. Può infatti essere che il nostro cervello non sia una macchina a stati discreti, ma a stati continui. Turing sottolinea però che l'interfaccia comunicativa di una macchina a stati continui non può che essere a stati discreti. In altre parole, anche se le macchine a stati continui fossero essenzialmente diverse rispetto a quelle a stati discreti, difficilmente questa differenza risulterebbe evidente all'interrogante[22].

Dal punto di vista della metodologia della psicologia, di cui abbiamo discusso nella sezione precedente, dove si colloca il gioco dell'imitazione? Di certo non tiene conto né della psicologia introspettiva, né della neuropsicologia. Turing non è interessato né alla coscienza, come abbiamo visto, né a come i neuroni influenzino il pensiero. O meglio, in questo articolo tiene fuori gioco tali questioni. D'acchito potrebbe sembrare che Turing sia un comportamentista, poiché in fondo stabilisce la presenza del pensiero mediante i comportamenti linguistici palesi. Come hanno mostrato Oppy e Dowe[23] non è così. Infatti il compito di programmare una macchina digitale capace di produrre quei comportamenti linguistici è estremamente complesso. In altre parole, Turing non sta immaginando gli stupidi arnesi proposti da Searle e Block, che non sono altro che degli enormi registri di stimoli e risposte; egli sta proponendo di mettere a punto un software altamente strutturato che si basi su un vero e proprio modello cognitivo. Possiamo quindi affermare che questo

21 - G. Oppy - D. Dowe, *The Turing Test*, in «The Stanford Encyclopedia of Philosophy» (Winter 2021 Edition), edited by Edward N. Zalta, in https://plato.stanford.edu/archives/win2021/entries/turing-test/, § 6.

22 - Va inoltre notato inoltre, che per quel poco che si sa sulle macchine a stati continui, come ad esempio lo Shannon's General Purpose Analog Computer (GPAC), sembra che siano in grado di calcolare esattamente come una macchina di Turing: O. Bournez - M.L. Campagnolo - D.S. Graça - E. Hainry, *Polynomial differential equations compute all real computable functions on computable compact intervals*. in «Journal of Complexity», 23 (3/2007), pp. 317-335.

23 - G. Oppy - D. Dowe, *The Turing Test*, cit.

saggio di Turing è una delle prime espressioni di quel movimento che ha dominato la psicologia della seconda metà del Novecento, che chiamiamo "cognitivismo".

Concludiamo questa sezione con le previsioni formulate da Turing. La prima è che nel 2000 sarà possibile programmare macchine che dopo 5 minuti del gioco dell'imitazione non vengano identificate nel 70% dei casi. Nella prossima sezione vedremo che di fatto questo non è successo. La seconda è che riuscire a superare il gioco dell'imitazione è soprattutto una questione di software e non di hardware. E anche su questo sappiamo che Turing si sbagliava, poiché le odierne reti neurali sono anche un importante progresso dal punto di vista ingegneristico. Tuttavia, anche se Turing non è del tutto esplicito al riguardo, il punto che l'intero articolo vuole sottolineare sembra essere un altro: le macchine digitali universali saranno sempre più brave a simulare il pensiero umano. E su questo, soprattutto gli ultimi dieci anni di ricerche tecnologiche hanno dato completamente ragione a Turing.

ChatGPT-4 ha superato il test di Turing?

Prima di entrare nel vivo della discussione del nostro tema, ancora alcune considerazioni generali.

Non è chiaro che cosa realmente rileverebbe il superamento del test di Turing. Come spesso capita in questo tipo di esperimenti, non possiamo certo dire che superare il test di Turing sia condizione necessaria e sufficiente per attribuire intelligenza o pensiero alle macchine. Neanche, però possiamo affermare che sia necessaria, poiché esistono senz'altro altre forme di intelligenza[24]. Neppure forse possiamo ritenere che sia sufficiente, in quanto si possono immaginare macchine del tutto stupide che potrebbero superarlo, come la "testa di legno" di Block o la "stanza cinese" di Searle. Questo argomento standard[25] non è però del tutto convincente, poiché questo tipo di macchine immaginate in esperimenti mentali filosofici è molto poco realistico. Come abbiamo già detto, un programma di computer in grado di parlare non può essere una lista enorme di input e di output, ma deve avere una sua struttura interna. Possiamo quindi dire che in linea di massima superare il test di Turing è una condizione sufficiente per attribuire a una macchina l'intelligenza. Resta il fatto, come abbiamo già sottolineato, che l'intelligenza di cui stiamo parlando è una nozione scientifica e cognitiva, che non ha molto a che fare con la nostra soggettività. Questa è stata la giusta scelta metodologica di Turing, che va rispettata.

Va inoltre sottolineato che, quando si realizza concretamente il test di Turing sono presenti molte variabili, che vanno opportunamente controllate. Di fatto stori-

24 - Vedi, ad esempio, G. Vallortigara, *Altre menti*, Il Mulino, Bologna 2022.

25 - G. Oppy - D. Dowe, *The Turing Test*, cit.

camente ci sono stati tentativi di somministrare il test, a partire da quello di ELIZA[26] a quelli legati al famoso premio istituito da Loebner[27]. In letteratura si trovano altri esperimenti, ma nessuno è conclusivo, fino ai due articoli che discuteremo di Jones e Bergen[28], che arrivano alla conclusione che ChatGPT-4 ha superato il test. Vediamo ora come è stato realizzato l'esperimento.

Nel primo articolo i due autori avevano provato diversi prompt. Nel secondo si concentrano su quello che meglio aveva performato, che è un prompt di ChatGPT-4. Non riporteremo tutti i dettagli dell'esperimento, solo quelli per noi rilevanti.

Una prima differenza essenziale rispetto alla situazione immaginata da Turing è che l'interrogante si trova di fronte a un solo "testimone", che può essere o ChatGPT-4 oppure una persona. L'interrogante, come proposto originariamente, ha a disposizione 5 minuti per indovinare se sta chattando con una persona o con una macchina. Alla persona-testimone vien chiesto di esprimersi come persona. L'interrogante ha graficamente di fronte una chat in cui scrive e ottiene risposte con un certo ritardo, anche quando sta messaggiando con una macchina, in modo da non notare la differenza. Il tasso di successo dell'interrogante viene misurato come il numero di volte che l'interrogante afferma di avere di fronte una persona. Il risultato più importante del primo esperimento è che nel caso in cui il testimone è un certo tipo di prompt di ChatGPT-4 su 855 casi l'interrogante ha sostenuto di avere davanti una persona il 50% delle volte. Per contro, quando aveva davanti una persona lo ha sostenuto il 66% delle volte. Non è immediato confrontare questo risultato con la previsione di Turing del 30% di identificazioni sbagliate. Teniamo presente che nell'esperimento proposto da Turing, il massimo che una macchina può ottenere è che l'interrogante sbagli il 50% delle volte, cioè che la sua ipotesi sia come il caso. Invece, nella situazione sperimentale di Jones e Bergen, che è diversa, il meglio per la macchina sarebbe performare come le persone, cioè essere scambiata per una persona il 66% delle volte. Facile allora fare la proporzione 50 : 66 = x : 50, cioè il tasso di successo dell'esperimento di Jones e Bergen è analogo al 38% in quello originale di Turing. Il test nella forma originale di Turing non è ancora stato realizzato, per cui questo confronto è solo virtuale. Tuttavia sembra che il test di Turing sia ampiamente superato. Si noti che ovviamente la dimestichezza con i *large language models* degli interroganti fa la differenza, nel senso che gli interroganti più esperti vengono ingannati di meno.

---

26 - J. Weizenbaum, *ELIZA – a computer program for the study of natural language communication between man and machine,* in «Communications of the ACM», 9 (1/1966), pp. 36-45.

27 - S.M. Shieber, *Lessons from a restricted Turing test*, in «*arXiv*» (1994) [cmp-lg/9404002].

28 - C.R. Jones - B.K. Bergen, *Does GPT-4 pass the Turing test?*, cit.; C.R. Jones - B.K. Bergen, *People cannot distinguish GPT-4 from a human in a Turing test*, cit.

Il secondo esperimento condotto da Jones e Bergen sostanzialmente conferma i dati del primo. Per questa ragione i due autori passano dal titolo interrogativo del primo articolo a quello affermativo del secondo. Su 500 testimoni del tipo migliore di prompt di ChatGPT-4 l'interrogante afferma di aver di fronte una persona – dopo una chat di 5 minuti – in più del 50% dei casi. Resta il fatto che le persone riescono a "sembrare" persone in una percentuale significativamente superiore di casi.

Tirando le somme, possiamo affermare che le previsioni di Turing si sono avverate, anche se con un certo ritardo. Le macchine digitali sono in grado di conversare con noi in un modo che offusca la nostra capacità di capire se stiamo dialogando con una macchina o con una persona. Al di là dell'entusiasmante risultato tecnologico, che con il prosieguo della ricerca non può che migliorare, il superamento del test di Turing pone problemi di carattere etico estremamente importanti. Infatti queste chatbot ormai quasi indistinguibili dalle persone possono essere usate in modi ingannevoli e dannosi per gli utenti.

Considerazioni conclusive

In *Machina sapiens*, Nello Cristianini usa un efficace apologo per descrivere la situazione che stiamo vivendo. È come se un'intelligenza aliena fosse atterrata fra di noi. Non aliena nel senso della sua origine, visto che la abbiamo messa a punto noi, ma nel senso che di fatto non sappiamo bene come funziona. ChatGPT era basata su 175 miliardi di parametri, cioè una rete neurale i cui pesi delle connessioni sono numeri fra 0 e 1 di 175 miliardi di nodi. Si suppone che ChatGPT-4 ne abbia almeno un ordine di grandezza in più. In pratica, anche se avessimo davanti a noi il tabulato di questi quasi 2000 miliardi di parametri e la loro assegnazione ai nodi della rete non sapremmo che cosa ChatGPT-4 stia facendo. Nel 1997, il calcolatore dedicato Deep Blue dell'IBM batté a scacchi il campione del mondo in carica Gary Kasparov. Fu un evento epocale. Gli ingegneri e i programmatori che avevano costruito Deep Blue, anche se non potevano seguire tutti i dettagli dei suoi processi, avevano ben chiaro come funzionasse la loro creatura. Essa aveva memorizzato migliaia di partite; considerata la posizione dei pezzi sulla scacchiera, analizzava la maggior parte degli sviluppi possibili per parecchie mosse e poi valutava in base a certi parametri decisi dai programmatori la qualità della posizione finale; sceglieva quindi la variante con il punteggio più alto. Per contro, il programma AlphaGo della DeepMind, che ha sconfitto 4 a 1 nel 2016 uno dei più forti giocatori del mondo di Go, cioè Lee Se-Dol, anche se ha una struttura di massima ben nota, nello scegliere le mosse del gioco ragiona in un modo sostanzialmente ignoto ai suoi costruttori.

Questa è una novità profonda nel mondo della tecnologia. Non sapendo bene come lavorano le reti neurali, non solo fatichiamo a modificarle, ma siamo di fronte a un notevole scacco morale per la nostra consapevolezza. Uno dei compiti che

ci aspetta è proprio quello di incrementare la nostra conoscenza di questi strumenti che ormai usiamo quotidianamente. Tanto più che essi, come abbiamo visto, hanno superato la barriera epocale del test di Turing. Questa maggiore consapevolezza può essere un problema informatico, nel senso di riuscire a trovare dei software che prendendo in input i parametri e la struttura di una rete neurale, diano in output una rappresentazione computazionale più comprensibile di come la rete stia operando. In questo senso nell'introduzione parlavamo della rivincita della logica. La logica, infatti, la capiamo, mentre le reti neurali no, quindi rappresentare logicamente le reti neurali è oggi un compito fondamentale.

Non dobbiamo però trascurare un altro aspetto interessante della comprensione delle reti neurali. Sean Trott e Timoty B. Lee, in questo bel post on line[29], spiegano in modo chiaro ed efficace come funzionano le reti neurali tipo ChatGPT. Non solo, nel breve testo ci sono anche i link alla maggior parte dei paper che nello spazio di pochi anni hanno portato nel 2017 alla prima messa a punto di ChatGPT. Dare un'occhiata a questi articoli è molto interessante, perché si entra nei meccanismi di lavoro delle teste pensanti che hanno messo a punto questa nuova tecnologia. Di fatto, sembra che questi ricercatori abbiano sviluppato strumenti linguistici naturali per descrivere e comprendere i *large language models* che stavano manipolando. In altre parole, essi stessi hanno elaborato dei parziali strumenti di comprensione che sono serviti a guidare le ricerche. Questi strumenti linguistici meritano la nostra attenzione e ulteriori ricerche, perché una loro disamina aiuterà ad attenuare il problema della comprensione delle reti neurali in cui oggi siamo immersi.

Ringrazio Marco Giunti e Massimo Mugnai, dai quali ho appreso molto riguardo a questi temi.

VINCENZO FANO
Professore ordinario di Logica e Filosofia della scienza presso il Dipartimento di Scienze Pure e Applicate, Università di Urbino
vincenzo.fano@uniurb.it

29 - [https://arstechnica.com/science/2023/07/a-jargon-free-explanation-of-how-ai-large-language-models-work/].

Agli albori dell'AI. Riflessioni sul test di Turing

di Francesco Federico Calemi

Over seventy years after its publication, Alan Turing's article Computing Machinery and Intelligence still fuels debate on the meaning and structure of the test it introduces, the limits of answering "Can machines think?", and the validity of that question itself. Widely seen as the first visionary manifesto on artificial intelligence, the article anticipated many ideas that became foundational in AI research. At a time when computers were mainly tools for high-speed calculation – first in military contexts and then in science – Turing used the concept of a computational machine to pose philosophical questions about human intelligence, the link between computation and thought, and the possibility of algorithmically modeling cognition. This essay reconstructs the test's structure and assumptions (§§ 1-3), explores the link between Turing's and Leibniz's views on thinking machines (§ 4), and examines major criticisms of the mental model in Turing's work (§ 5).

Keywords: Turing Machine; Functionalism; Imitation Game; Wittgenstein; Gödel

L'intuizione di Turing

L'elaborazione del «gioco dell'imitazione» — noto come *test di Turing*, in omaggio al suo ideatore — avvenne in un periodo storico vicino alla Seconda Guerra Mondiale e ancora profondamente segnato dai suoi tragici avvenimenti. Proprio tra il 1940 e la conclusione del conflitto, Alan Turing, lavorando segretamente nell'unità di controspionaggio britannico riunita a Bletchley Park, contribuì in modo determinante allo sviluppo di *The Bombe*, un dispositivo concepito per decifrare i codici nazisti generati dalla macchina *Enigma*[1]. Come scrive Copeland:

> La guerra fu una parentesi disastrosa per molti, ma per Turing rappresentò uno sbocco di importanza internazionale per il suo genio creativo. Non è esagerato affermare che, senza Turing, la guerra sarebbe probabilmente durata più a lungo e avrebbe potuto essere vinta dai nazisti[2].

Infatti, l'uso dei calcolatori in ambito bellico rivelò l'enorme potenziale rappresentato dalle macchine nel gestire compiti complessi e stratificati, come il ragionamento logico e la ricerca euristica, attività tipicamente associate all'intelligenza umana. Fu proprio nel corso di questo conflitto devastante che la ricerca scientifico-tecnologica compì progressi straordinari nell'ambito della computazione. In particolare, secondo una testimonianza raccolta da Jack Copeland in un'intervista a Donald Michie, collega di Turing a Bletchley Park, sembra che già in quegli anni Turing stesse maturando un'intuizione di vasta portata[3]: che i calcolatori non solo potessero essere programmati per risolvere problemi tecnici specifici, ma che, in linea di principio, avrebbero potuto simulare processi cognitivi umani, gettando le basi per quella che sarebbe poi divenuta la moderna intelligenza artificiale (AI), rispetto alla quale il test di Turing rappresenta un passo fondamentale e preliminare. Cercheremo, in quanto segue, di chiarire in cosa consista il test e quale sia il suo significato logico-filosofico.

---

1 - Rinviamo per maggiori approfondimenti alla raccolta J. Copeland *et al.* (a cura di), *The Turing Guide*, Oxford University Press, Oxford 2017.

2 - *Ibid.*, p. ix.

3 - «After a heavy week of codebreaking, the two would meet in a village pub on Friday evenings to discuss how to reproduce human thought processes in a universal Turing machine […] From chess Turing moved on to the idea of automating the whole process of learning» (J. Copeland, *Intelligent Machinery*, in J. Copeland *et al.* [a cura di], *The Turing Guide*, cit., p. 267).

Il gioco dell'imitazione

Come è noto, Turing dedica la prima sezione dell'articolo *Computing Machinery and Intelligence*[4] alla formulazione (e alla successiva ri-formulazione) di quella che è la domanda-guida dell'intero scritto: «le macchine possono pensare?». La provocatorietà di tale interrogativo potrebbe suggerire una lettura semplificata e ingenua secondo la quale il test mostrerebbe che le macchine, in qualche modo, possano pensare esattamente nello stesso senso in cui diciamo che gli esseri umani pensano. Senza attribuire a questa lettura un effettivo seguito in ambito filosofico, la considereremo come una posizione limite, utile a mettere in luce – per contrasto – il corretto quadro teorico entro cui si può cogliere il significato autentico del test di Turing.

Anzitutto lo stesso Autore precisa che per rispondere in modo diretto al quesito «le macchine possono pensare?» occorre un preliminare chiarimento dei termini "macchina" e "pensare". Tuttavia, osserva Turing, ciò non può avvenire imponendo delle definizioni, né cercando di comprendere il modo in cui adoperiamo tali espressioni nelle nostre pratiche linguistiche quotidiane. Nel primo caso si rischierebbe, infatti, di predeterminare arbitrariamente una risposta, mentre nel secondo caso il tentativo di fornire una risposta richiederebbe, assurdamente, una indagine demoscopica. Turing, a fronte di queste difficoltà, propone una peculiare riformulazione della domanda iniziale per mezzo dell'espediente del «gioco dell'imitazione»[5], di cui è opportuno esporre sinteticamente set up, scopo e svolgimento.

Il gioco dell'imitazione prevede tre giocatori, un uomo (giocatore A), una donna (giocatore B) e un interrogante, situati in stanze differenti e reciprocamente isolate in modo che nessuno dei tre possa avere accesso alle altre stanze. (Questa condizione di inaccessibilità fisica risulterà, come vedremo, di estrema importanza per comprendere il problema dell'attribuzione di stati mentali.) L'interrogante può tuttavia comunicare con l'uomo e la donna tramite una telescrivente, mentre non è previsto che l'uomo e la donna comunichino in alcun modo tra di loro, né che in qualche modo possano dare prova effettiva di quanto di volta in volta dichiarano.

Le condizioni di vittoria del gioco sono le seguenti: lo scopo dell'interrogante è determinare chi è l'uomo e chi è la donna solo per mezzo dell'intercorsa comunicazione scritta. Lo scopo del giocatore B è quello di aiutare l'interrogante ad effettuare una corretta identificazione, mentre lo scopo del giocatore A è quello di depistare l'interrogante impedendogli di effettuare una corretta identificazione.

---

4 - A. Turing, *Computing Machinery and Intelligence*, in «Mind», 59 (1950), pp. 433-460. Nel presente lavoro faremo riferimento alla seguente traduzione italiana: A. Turing. *Meccanismo computazionale e intelligenza artificiale*, tr. it. a cura di L. Cimmino, Città Nuova, Roma 2018.

5 - Turing, *Meccanismo computazionale e intelligenza artificiale*, cit., pp. 7-8.

Per dirlo con le parole di Turing, per B «la migliore strategia potrebbe essere quella di fornire risposte vere»[6] laddove la migliore strategia di A è quella di far sì che l'interrogante «sbagli l'identificazione»[7].

Naturalmente per l'interrogante risulta altamente difficoltoso vincere al gioco dell'imitazione dato che i soli dati di cui dispone per poter effettuare le proprie identificazioni sono di tipo linguistico: nessuna ispezione visiva, tattile o comunque sensoriale gli è permessa nel tentativo di indicare, tra i due giocatori coi quali comunica con telescrivente, chi è l'uomo e chi è la donna.

Definito il gioco dell'imitazione, Turing chiede quanto segue:

> "Cosa accade se è una macchina ad assumere il posto di A nel gioco?". L'interrogante sbaglierà tanto spesso quanto avverrebbe se a giocare fossero un uomo e una donna?[8].

Turing aggiunge: «Queste domande sostituiscono la domanda iniziale "Le macchine possono pensare?"»[9]. In altri termini, la domanda-guida dello scritto di Turing diviene la seguente: un essere umano può ingannarsi nell'identificare e riconoscere, a partire dalla sola ispezione del comportamento linguistico, una macchina e un essere umano? Questa riformulazione rende chiaro che il focus della domanda non è il (presunto) soggetto-macchina e la sua (presunta) capacità di avere stati mentali, ma diviene propriamente il soggetto umano: quando e in quali condizioni attribuiamo l'attività del pensare? In quali condizioni giudichiamo *intelligente* un comportamento?

Naturalmente, nel quadro dell'esperimento mentale considerato, si dirà che una macchina "supera" il test di Turing se l'interrogante sbaglierà con sufficiente frequenza le sue identificazioni. Su questa possibilità Turing si esprime nel testo in senso pienamente positivo sostenendo quanto segue:

> Credo che tra cinquant'anni sarà possibile programmare i computer in grado di partecipare al gioco dell'imitazione così bene che un interrogante medio non avrà più del 70% di possibilità di operare la corretta identificazione dopo cinque minuti di domande[10].

Per quanto Turing si mostrasse fiducioso che tale soglia sarebbe stata raggiunta in tempi brevi, è solo nei primissimi anni Venti del XXI secolo che alcuni modelli linguistici avanzati – come GPT-4 – hanno cominciato a superare versioni rigorose del test, sia pure con circa venticinque anni di ritardo. È tuttavia importante preci-

---

6 - *Ibid.*, p. 8.

7 - *Ibidem.*

8 - *Ibidem.*

9 - *Ibidem.*

10 - *Ibid.*, pp. 19-20.

sare che si tratta di un traguardo solo parziale: tali sistemi, per quanto sofisticati nella simulazione del linguaggio umano, non possiedono ancora una vera intelligenza artificiale generale, e la loro "comprensione" resta confinata a domini specifici e a comportamenti appresi statisticamente. Tuttavia, ciò su cui focalizzeremo l'attenzione in questo contesto non è l'effettiva implementazione storica del test, ma la sua struttura concettuale e il tipo di attribuzioni mentali che esso implica.

La macchina universale di Turing

La seconda questione da chiarire per comprendere il significato del test di Turing è quella riguardante le macchine coinvolte nel gioco dell'imitazione. Le macchine a cui si fa riferimento – che qui denomineremo come di consueto *Macchine di Turing* (MT) – sono dispositivi progettati in stretta analogia con il modo in cui un essere umano svolge i propri calcoli, ossia in ottemperanza a regole operative specificate in un manuale o memorizzate, e svolgendo calcoli su fogli di carta (ipoteticamente disponibili in quantità illimitata). Per tale motivo una MT è composta da un'unità esecutiva, un'unità di controllo, e un dispositivo di archiviazione. In termini pratici, la macchina dispone di un nastro mobile, potenzialmente illimitato, suddiviso in riquadri su ciascuno dei quali è possibile stampare un solo simbolo alla volta, di un lettore ottico e un dispositivo di stampa. Ogni stato della macchina sarà pertanto determinato da una quintupla ordinata costituita da, rispettivamente, la configurazione iniziale della macchina in un dato momento, *t*, il simbolo scansionato in *t*, l'operazione da svolgere, il movimento del nastro, e la configurazione finale assunta dalla macchina[11].

Successivamente Turing generalizza la nozione di MT mettendo a punto il modello di una macchina universale, che in questa sede denomineremo *macchina di Turing universale* (MTU). Una MTU è in grado, infatti, di ricevere come input sul proprio dispositivo di archiviazione una rappresentazione aritmetizzata delle istruzioni operative di qualsiasi altra MT e di simularne perfettamente il comportamento[12]. Turing identifica quindi una MTU con un *computer digitale* ossia una macchina a stati-discreti tale da "saltare" da uno stato interno all'altro, e tale da rispondere a determinati input con adeguati output, imitando all'occorrenza, con un'apposita

11 - Per una descrizione più accurata si veda M. Davis. *Il calcolatore universale. Da Leibniz a Turing*, tr. it. A cura di G. Lolli, Einaudi, Torino 2012 e F.F. Calemi, *Macchine e pensiero. Assiomatizzazione e limiti della computabilità a partire da Turing*, in M. Falcone - E. Pili (a cura di), *L'irriducibile umano. Etica e diritto delle intelligenze artificiali*, Il Mulino, in corso di pubblicazione.

12 - Per un approfondimento rinviamo al saggio introduttivo di Gabriele Lolli in A. Turing, *Intelligenza meccanica*, tr. it. a cura di G. Lolli, Bollati Boringheri, Torino 1992.

programmazione, il comportamento di qualsiasi altra macchina a stati-discreti. Come rileva Copeland:

> leggendo ed eseguendo le istruzioni che il programmatore memorizza sul nastro, la macchina di Turing universale [...] è in grado di eseguire qualsiasi computazione possibile[13].

Turing è fermamente convinto che, poiché l'attività del pensare è legata al funzionamento del cervello umano e, in particolare, poiché il cervello umano può essere interpretato come una macchina a stati discreti (su questo punto cruciale dovremo tornare nel prosieguo), un computer digitale, inteso come una MTU, possa simularne il funzionamento. Di conseguenza, se il soggetto interrogante non riesce a superare il test di Turing, è ragionevole considerare "pensiero" anche il prodotto dell'attività di un computatore digitale, per quanto controintuivito ciò possa sembrare.

Per comprendere più a fondo tale prospettiva riteniamo utile proporre un parallelismo tra l'argomento di Turing e quello sviluppato da Leibniz nei *Principi della filosofia* in merito al problema delle "macchine pensanti": ciò aiuterà a comprendere l'innovatività della proposta di Turing in merito al dibattito sulla natura del mentale.

Dal mulino pensante al computatore digitale

In un celebre passo dei *Principi della filosofia* Leibniz invita il lettore ad immaginare, come parte di una *reductio ad absurdum* di ogni spiegazione meccanicistica del pensiero, di poter visitare e osservare dall'interno una presunta "macchina pensante", similmente a come si potrebbe entrare in un mulino osservandolo internamente. Con questa vivida immagine Leibniz sostiene che, anche scomponendo meccanicamente il cervello, non si sarebbe trovato nulla in grado di spiegare il pensiero, ma «solo parti che spingono altre parti»:

> D'altra parte si deve riconoscere che la percezione, e quel che ne dipende, *è inesplicabile mediante ragioni meccaniche*, cioè mediante le figure e i movimenti. Immaginiamo una macchina strutturata in modo tale che sia capace di pensare, di sentire, di avere percezioni; supponiamola ora ingrandita, con le stesse proporzioni, in modo che vi si possa entrare come in un mulino. Fatto ciò, visitando la macchina al suo interno, troveremo sempre e soltanto pezzi che si spingono a vicenda, ma nulla che sia in grado di spiegare una percezione. Quindi la [ragio-

13 - J. Copeland, *Turing's great invention: the universal computing machine*, in J. Copeland *et al.* (a cura di), *The Turing Guide*, cit., p. 54.

> ne della] percezione va cercata nella sostanza semplice, non già nel Composto, cioè nella macchina[14].

Ogni macchina, dunque, per quanto complessa, rimane un'entità puramente materiale, composta da parti meccaniche che si muovono secondo leggi fisiche prestabilite. Di conseguenza, nell'ipotesi che l'attività mentale sia semplicemente il prodotto di una macchina, e ipotizzando di poter estendere la scala della macchina fino a renderla esplorabile, troveremmo al suo interno solo movimenti e meccanismi, ma mai la *percezione*, la *coscienza*, o, in generale, il *pensiero*. Ne deriva che ogni tentativo di spiegare la mente esclusivamente in termini materialistici o meccanicistici risulta destinato al fallimento.

La posizione leibniziana potrebbe dunque essere riassunta dal seguente argomento: se la mente fosse una macchina, allora sarebbe una macchina pensante; se la mente fosse una macchina pensante, allora in essa vi sarebbero pensieri (percezioni, coscienza); ma in una macchina vi sono solo e soltanto parti meccaniche; dunque la mente non è una macchina. Non ci addentreremo nei dettagli dell'argomento di Leibniz[15], ma considereremo l'originalità dell'approccio al problema del mentale che emerge entro il contesto offerto da Turing e dal suo test.

Anzitutto occorre sottolineare che Turing legge l'analogia mente-macchina in una prospettiva funzionalista evitando il binomio materiale-immateriale: l'implementabilità della MTU mostra che il pensiero può essere descritto in termini di calcolo simbolico, rendendo la meccanicità del calcolo non una negazione del pensiero stesso, ma una sua rappresentazione in termini funzionali. Inoltre, e proprio per via di questa impostazione, Turing evita il problema della riducibilità degli stati mentali posto da Leibniz, mostrando che l'autentica questione non è il tentare di spiegare la natura del pensiero col "collocarlo" in uno spazio non-fisico (quello del mentale, per l'appunto), ma nella comprensione dei processi astratti che un supporto fisico può realizzare. Da ciò consegue che Turing, concentrandosi su ciò che una macchina può fare e su come il suo comportamento possa essere indistinguibile da quello umano, non cerca precipuamente di indagare la natura ontologica della mente, ma sposta l'attenzione sulla predisposizione umana ad attribuire (o negare) attività pensante in prossimità di comportamenti linguistici osservabili. Ciò consente di sostenere che il test di Turing ritrova il suo autentico scopo e significato nel quadro di una valutazione di quanto e come una macchina possa replicare comportamenti intelligenti *indipendentemente dalla loro* (*eventuale*) *natura intrin-*

---

14 - G. W. Leibniz, *Principi della filosofia o Monadologia*, trad. it. di S. Cariati, Bompiani, Milano 2001, p. 65.

15 - Per una discussione più approfondita rinviamo a P. van Inwagen, *Metafisica,* tr. it. a cura di F.F. Calemi, Cantagalli, Siena 2011, pp. 325-365.

seca. Questa posizione è stata successivamente battezzata come *Turing machine functionalism*, e risulta sintetizzabile, per usare le parole di Dennett, nella seguente maniera:

> per ogni predicato "mentale" "M" esiste un predicato "F" esprimibile in un linguaggio fisicamente neutro, ma progettato per specificare funzioni astratte e relazioni funzionali. I candidati più ovvi per un tale linguaggio sono i sistemi utilizzati per descrivere i computer o i programmi[16].

Nella riflessione sulla filosofia della mente, infatti, il funzionalismo è quella posizione che non richiede la traducibilità sistematica tra predicati mentali e predicati fisicamente neutri. Per il funzionalismo la natura di uno stato mentale non dipende dalla sua struttura interna, ma dalla funzione che svolge e dal ruolo che ricopre all'interno del sistema di cui fa parte:

> Turing identifica i pensieri con stati di un sistema determinati esclusivamente dal loro ruolo nel generare ulteriori stati interni e risposte verbali, a partire da determinati input linguistici. Secondo questa prospettiva [...] diversi sistemi, anche fisicamente molto diversi tra loro, possono avere stati interni [qualitativamente differenti] che svolgono le medesime funzioni[17].

Pertanto sia lo stato interno di una MTU, sia lo stato interno di un computatore umano (costituito da una certa attivazione del suo sistema nervoso), possono svolgere, nonostante l'enorme differenza qualitativa che li contraddistingue, la medesima funzione; e se attribuiamo intelligenza all'output di tale funzione, allora contraddistinto da intelligenza sarà anche l'operatore che l'ha prodotto, chiunque *egli/esso* sia.

Critiche al modello di Turing

Le considerazioni sviluppate da Turing hanno ispirato numerose riflessioni sulla adeguatezza dell'analogia mente-calcolatore. Notoria la posizione di Putnam, tra le principali, che in un'opera del 1961 radicalizza il funzionalismo di Turing:

> Vorrei suggerire che ci sono molte considerazioni che puntano all'idea che una macchina di Turing più elementi casuali sia un modello ragionevole per il cervello umano[18].

---

16 - D. Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology*, Bradford Books, Cambridge 1978, p. XV.

17 - J. Levin, *Functionalism*, in E.N. Zalta - U. Nodelman (a cura di), *The Stanford Encyclopedia of Philosophy*, Summer 2023, Metaphysics Research Lab, Stanford University, 2023.

18 - H. Putnam, *Some issues in the theory of grammar*, in H. Putnam, *Mind, Language and Reality: Philosophical Papers, volume 2*, Cambridge University Press, Cambridge 1975, p. 102.

Tuttavia le critiche che la posizione di Turing ha incontrato sono svariate. Uno dei presupposti cruciali dell'argomento di Turing è che il cervello umano possa essere descritto come un meccanismo a stati-discreti. Come già detto, un meccanismo a stati discreti è un sistema che opera attraverso un numero finito di stati ben definiti, con transizioni tra stati determinate da regole precise e input specifici. Turing afferma che

> Il sistema nervoso non è certamente una macchina a stati-discreti. Un piccolo errore d'informazione sull'intensità di un impulso nervoso che colpisce un neurone può causare un enorme differenza sull'intensità dell'impulso in uscita[19].

Tuttavia Turing appare persuaso che, nonostante la sua apparente complessità biologica e la sua presunta struttura continua, il funzionamento del sistema nervoso possa essere descritto nei termini di un modello finitistico e discreto, in cui segnali e processi sono rappresentabili attraverso stati distinti e regole di transizione[20]. Ma se la fisiologia del sistema nervoso può essere descritta nei termini di una macchina a stati-discreti, e se una MTU è in grado di emulare il comportamento di qualsiasi macchina a stati-discreti, allora una MTU può emulare l'attività del cervello umano. E, come visto, sotto le specifiche condizioni previste dall'esperimento mentale del gioco dell'imitazione, sarebbe perfino possibile attribuire "pensieri" o stati mentali a un dispositivo che, di per sé, non possiede la fisiologia del cervello umano.

Questo punto, come già anticipato, è certamente uno dei più delicati della strategia argomentativa di Turing e indubbiamente anche uno dei più contestati dai suoi critici. Gödel, ad esempio, da un lato non esita a riconoscere i meriti delle implicazioni logico-matematiche dell'opera di Turing, scrivendo quanto segue nella celebre *Gibbs Lecture* del 1951:

> Ci sono vari percorsi diversi per arrivare a questa definizione [di procedimento finito]. Il più soddisfacente, a mio parere, è la riduzione del concetto di procedimento finito a quello di una macchina con un numero finito di componenti, come è stato fatto dal matematico inglese Turing[21]

---

19 - Turing, *Meccanismo computazionale e intelligenza artificiale*, cit., p. 32.

20 - Precisiamo che sebbene Turing sembri assumere che il sistema nervoso operi in maniera continua e non discreta, questa ipotesi non è affatto scontata. Esistono anzi buone ragioni per considerare che, almeno a un certo livello di astrazione, il comportamento neuronale possa essere descritto efficacemente come un sistema a stati discreti. Alcuni modelli di neuroscienze computazionali – così come l'intera architettura delle reti neurali artificiali – si fondano proprio su tale ipotesi.

21 - K. Gödel, *Alcuni teoremi basilari sui fondamenti della matematica e loro implicazioni filosofiche*, in K. Gödel, *Opere. Saggi inediti e conferenze*, tr. it. a cura di G. Lolli - C. Mangione - E. Ballo, Vol. 3, Bollati Boringheri, Torino 2002, p. 268

ed esprimendo, nel passo successivo, forti riserve sulle applicazioni filosofiche di tali innovazioni:

> Per quel che concerne le conseguenze filosofiche dei risultati che esaminiamo, io non credo che siano mai state discusse adeguatamente, o semplicemente [siano state] notate[22].

Per Gödel infatti le teorie riduzioniste della mente trascurano una disgiunzione fondamentale:

> o la mente umana (perfino nell'ambito della matematica pura) supera infinitamente la potenza di ogni macchina finita, oppure esistono proposizioni matematiche assolutamente indecidibili, ossia verità che sfuggono non solo a ogni sistema formale, ma anche a ogni possibile forma di intuizione matematica umana[23].

Tale premessa ha, per Gödel, una conseguenza essa stessa disgiuntiva, ossia o

> il funzionamento della mente umana non può essere ridotto al funzionamento del cervello che secondo ogni apparenza è una macchina finita con un numero finito di componenti, cioè i neuroni e i loro collegamenti[24]

oppure «esistono proposizioni matematiche assolutamente indecidibili» e ciò porterebbe a negare «l'idea che la matematica sia soltanto una nostra creazione»[25]. Benché nel 1951 Gödel non scelga esplicitamente tra le due alternative, mostra comunque una preferenza per la prima, suggerendo che la mente umana possa accedere a verità che nessuna macchina è in grado di generare. Nel 1972 la sua posizione diviene meno sfumata, giungendo a sostenere che

> Turing [...] propone un'argomentazione che si ritiene mostri come i processi mentali non possano andare al di là di quelli meccanici. Tuttavia l'argomentazione non è conclusiva[26].

La preoccupazione di Gödel, infatti, risiede nel fatto che il modello della mente proposto da Turing manifesta un limite fondamentale: l'incapacità di catturare il carattere dinamico e in evoluzione del pensiero umano. Secondo Gödel, invece, la mente non è un sistema statico che opera con un insieme fisso di concetti o strumenti, ma un'entità che si sviluppa costantemente, ampliando sia il numero di termini astratti che è in grado di comprendere sia la precisione con cui questi vengono

---

22 - *Ibid.*, p. 274.

23 - *Ibidem*.

24 - *Ibid.*, p. 275.

25 - *Ibidem*.

26 - K. Gödel. *Alcune osservazioni sui risultati di indecidibilità*, in K. Gödel, *Opere. 1938-1974*. tr. it. a cura di G. Lolli - C. Mangione - E. Ballo, Vol. 2, Bollati Boringheri, Torino 2002, p. 306.

intesi e definiti. Questo processo evolutivo potrebbe essere guidato da metodi sistematici, ma anche in presenza di vincoli finiti in ogni momento, la sua crescita cumulativa può tendere all'infinito:

> anche se a ogni stadio il numero e la precisione dei termini astratti a nostra disposizione può solo essere finito, entrambi (e quindi anche il numero degli stati distinguibili della mente introdotti da Turing) possono nel corso dell'applicazione del processo convergere all'infinito[27].

In sostanza, Gödel sottolinea che la mente ha una capacità potenzialmente illimitata di espandere il proprio orizzonte concettuale, andando oltre i confini rigidi di un modello puramente meccanico come quello immaginato da Turing e codificato dalla MTU.

Analogamente, Nagel e Newman sottolineano, in relazione ai teoremi di incompletezza di Gödel e con un chiaro riferimento implicito a Turing, che

> Il cervello umano [...] sembra incarnare una struttura di regole operative che è decisamente più potente rispetto alla struttura delle macchine artificiali concepite fino ad oggi [...]. La dimostrazione di Gödel [...] implica che le risorse dell'intelletto umano non sono state e non possono essere completamente formalizzate, e che nuovi principi di dimostrazione attendono costantemente di essere inventati e scoperti [...]. Il teorema indica che la struttura e le capacità della mente umana sono molto più complesse e sottili rispetto a qualsiasi macchina non vivente immaginata finora[28].

Non meno importante, in tale quadro, risulta la critica sollevata dal *Chinese Room Argument* di Searle, volta a minare l'idea che un sistema computazionale possa conseguire una vera comprensione del significato delle informazioni che elabora. Searle utilizza il caso ipotetico di una persona che si trova in una stanza e manipola con maestria simboli costituiti da ideogrammi cinesi senza comprenderne il significato, per mostrare che la mera manipolazione di simboli, operata in ottemperanza a manuali di traduzione meramente formali, non implica necessariamente la possibilità di attribuire pensiero o comprensione all'operatore che svolge il processo di manipolazione simbolica[29].

---

27 - *Ibid.*, p. 306.

28 - J. Newman - E. Nagel, *Gödel's Proof*, Routledge, New York 1958, pp. 100-101. Appare peraltro, anche alla luce di quanto indicato fin qui, che la posizione di Nagel e Newman sorvola – o quantomeno lascia sullo sfondo – le considerazioni svolte da Gödel nella *Gibbs lecture* del 1951 sulla natura disgiuntiva delle conseguenze filosofiche dei risultati di incompletezza.

29 - Per un approfondimento rimandiamo a F.F. Calemi, *Significato, pensiero e soggettività: origine e sviluppi della filosofia analitica*, in M. Marianelli *et al.* (a cura di), *Anima, corpo, relazioni. Storia della filosofia da una prospettiva antropologica*, Vol. 3, Città Nuova, Roma 2022, pp. 467-480.

Ricordiamo infine una critica ante-litteram proposta da Wittgenstein sul tema delle macchine pensanti, che anticipa alcune delle questioni sollevate venti anni dopo da Turing. Poco si ricorda, infatti, che la domanda «Una macchina può pensare?» venne già sollevata da Wittgenstein negli appunti che nel 1933-34 dettò ai propri studenti a Cambridge, poi raccolti nel testo pubblicato postumo noto come *Libro blu*. Wittgenstein, nel definire il pensiero come un *operare con segni*[30], anticipa una possibile obiezione: «Quest'obiezione è espressa nella domanda: "Potrebbe una macchina pensare?"»[31]. Come spesso accade nei testi di Wittgenstein scritti dopo il *Tractatus*, a domande si tenta di rispondere sollevando altre domande che mettano in luce l'insensatezza delle prime: «Per ora nominerò soltanto una domanda analoga: "Può una macchina avere il mal di denti?"». Wittgenstein evidenzia che la vera difficoltà del quesito non sta nel fatto che non abbiamo ancora costruito una macchina in grado di "pensare", ma nel fatto che l'usuale grammatica della parola "pensare" viene violata dalla stessa posizione della domanda: quello del *pensare* sarebbe in altri termini un gioco linguistico a cui la macchina non può prendere parte. Nelle parole di Wittgenstein:

> Il grattacapo consiste […] nel fatto che la frase "una macchina pensa" (percepisce, desidera) sembra in qualche modo priva di senso. È come se avessimo chiesto: "Il numero 3 ha un colore?". ("Che colore potrebbe essere, visto che ovviamente non ha nessuno dei colori a noi noti?")[32].

Benché spesso i tentativi di affrancamento da tali critiche si basino sul sottolineare che il motivo dominante dell'opera di Turing sia l'attribuibilità di intelligenza a sistemi computazionali operanti nei contesti specifici previsti dal gioco dell'imitazione, resta il fatto che, sullo sfondo di tale riflessione e del parallelismo funzionale da essa implicato, permanga l'idea della riducibilità computazionale dell'intelligenza umana. Questa concezione, sebbene abbia avuto sostenitori nell'ambito

---

30 - Per motivi di spazio non possiamo approfondire la questione in questa sede e pertanto rinviamo a F.F. Calemi, *Parlare «come gioca uno scacchista». Matematica e semantica degli scacchi in Wittgenstein*, in M. Moschini (a cura di), *Il riconoscimento in gioco*, Pièdimosca, Perugia 2021, pp. 33-53.

31 - L. Wittgenstein, *Libro blu e libro marrone*, Einaudi, Torino 2000, p. 25.

32 - *Ibid.*, pp. 65-66. Si noti come la posizione di Wittgenstein in merito alla domanda «Potrebbe una macchina pensare?» è esattamente la medesima che Turing sposa nel suo articolo 1950: «Penso che la domanda iniziale, "Le macchine possono pensare?" sia tanto priva di significato da non meritare di essere discussa» (A. Turing, *Meccanismo computazionale e intelligenza artificiale*, cit., pp. 19-20). Tuttavia, mentre in Wittgenstein vi è l'intento terapeutico di dissolvere il «grattacapo» filosofico mostrando l'inadeguatezza dell'uso linguistico fuori dal contesto delle forme di vita *umane*, Turing aggira il riferimento a tali forme di vita attraverso una riformulazione operativa del quesito, motivata dall'intento di esplorare le potenzialità del modello rappresentato dalla MTU.

dell'intelligenza artificiale e della filosofia della mente[33], solleva questioni di notevole complessità, sia dal punto di vista epistemologico che ontologico. Infatti, da un lato, la possibilità di modellare il pensiero umano attraverso sistemi computazionali impone di chiarire in che misura la nozione di computazione possa davvero esaurire la ricchezza dei processi cognitivi; dall'altro, la prospettiva riduzionista si scontra con le difficoltà teoriche legate alla coscienza, alla comprensione semantica e alla natura dell'intenzionalità. Come appena visto, la questione rimane altamente problematica e ancora aperta, lasciando spazio a ulteriori indagini e a possibili sviluppi teorici che potrebbero ridefinire i confini tra mente e macchina[34].

Conclusione

Al di là delle aporie da essa sollevate, resta il fatto che la riflessione di Alan Turing rappresenta un punto di svolta nella storia del pensiero scientifico e filosofico, offrendo una nuova prospettiva sulla relazione tra mente, macchina e intelligenza. Il test di Turing rappresenta, come visto, non una mera verifica da sottoporre a un computer ma un invito a riflettere criticamente sul modo in cui attribuiamo significato e valore al comportamento intelligente: la sua eredità si estende ben oltre il campo dell'informatica, ispirando discussioni interdisciplinari che abbracciano filosofia, psicologia, neuroscienze e scienze cognitive.

L'opera di Turing ha quindi aperto la strada a una comprensione più critica e metodologicamente consapevole della natura del pensiero e dell'intelligenza, rappresentando – sia come modello a cui ispirarsi sia come posizione da cui prendere le distanze – un punto di riferimento per ogni interrogazione, oggi più che mai attuale, sul rapporto tra uomo e macchina, natura ed emulazione, pensiero e calcolo[35].

FRANCESCO F. CALEMI
Professore associato di Logica e Filosofia della logica presso il Dipartimento di Filosofia, Scienze sociali, Umane e della Formazione dell'Università degli studi di Perugia
francesco.calemi@unipg.it

33 - Come, ad esempio, W. McCulloch - W. Pitts, *A Logical Calculus of the Ideas Immanent in Nervous Activity*, in «Bulletin of Mathematical Biophysics», 7 (1943), pp. 115-133; H. Putnam, *Some issues in the theory of grammar*, cit.

34 - Per una rassegna puntuale delle principali posizioni che criticano il modello mentale computazionale si veda L. Cimmino, *Test di Turing, intenzionalità e solipsismo*, in A. Turing, *Meccanismo computazionale e intelligenza artificiale*, cit., pp. 51-178.

35 - L'autore desidera ringraziare un referee anonimo per i preziosi commenti forniti su una versione precedente del presente articolo.

Virtù Intellettuali e Intelligenza Artificiale: Il Machine Learning può fornire una Comprensione Scientifica del mondo?

di Vera Matarese

The pervasive use of Artificial Intelligence (AI) in the contemporary scientific method raises a crucial question: can AI provide a scientific understanding of the world? In this paper, we will answer this question by interpreting the concept of scientific understanding as an intellectual virtue inherent to the person. This interpretation is fruitful, as it allows us to transcend the conceptual horizons of science, offering a fertile ground for discussion for all those areas in which AI is used for epistemic purposes.

Keywords: Machine Learning; Intellectual Virtues; Scientific Understanding; Interpretability; Modal relations

Introduzione

L'intelligenza artificiale (IA) rappresenta oggi una componente essenziale e pervasiva del metodo scientifico, intervenendo in ogni fase del processo di ricerca scientifica[1]. L'IA viene impiegata per raccogliere, generare e interpretare i dati, per formulare ipotesi che possano spiegarli e selezionare quella più probabile. Viene anche utilizzata per progettare gli esperimenti atti a verificare le ipotesi scientifiche indagate e per ottimizzare i processi sperimentali stessi, ad esempio trovando la combinazione di parametri sperimentali che sia ottimale per la riuscita dell'esperimento. L'IA si sta affermando come pilastro imprescindibile anche per il suo ruolo rivoluzionario nel processo di predizione e di scoperta scientifica, come dimostrano casi emblematici quali l'individuazione di nuovi esopianeti[2] o la sintesi di nuovi nanomateriali[3]. Le sue capacità, che sono trasversali a ogni ambito disciplinare scientifico, stanno rivoluzionando il *modus operandi* della scienza. I benefici sono innegabili. Predicendo quali sono le ipotesi più promettenti e affinando in modo autonomo le condizioni sperimentali rendendole ottimali, l'IA consente un'ottimizzazione significativa di tempi e costi, imprimendo un'accelerazione senza precedenti al progresso scientifico. È importante sottolineare che questa rivoluzione non si esaurisce nella dimensione operativa, ma investe profondamente anche la sfera epistemica. Anzitutto, l'IA sta emergendo come un autentico *peer* epistemico, un collaboratore scientifico a tutti gli effetti, al punto che molte scoperte iniziano a essere quasi attribuite a essa. Inoltre, si osserva che l'uso dell'IA consente alla scienza di ottenere progressi epistemici notevoli, sia in campo predittivo sia concernenti la scoperta di nuove realtà, e che però questi, non necessariamente, si accompagnano a un progresso nella comprensione dei sistemi fisici indagati[4]. A proposito di questo, la diffusione capillare dell'IA solleva una questione cruciale: può l'IA fornire una comprensione scientifica dei sistemi fisici? Se la risposta a questa domanda dovesse essere dubbia, o negativa, il rischio è che la diffusione capillare dell'IA potrebbe portare a uno slittamento dell'asse epistemica della scienza verso il suo carattere predittivo piuttosto che quello propriamente conoscitivo che riguarda la comprensione del mondo. L'interrogativo se l'IA possa fornire

1 - Per una discussione sui ruoli dell'IA nella ricerca scientifica, si veda: H. Wang et al., *Scientific discovery in the age of artificial intelligence*, in «Nature», 620-7972 (2023), pp. 47-60.

2 - A. Valizadegan, et al., *ExoMiner: A highly accurate and explainable deep learning classifier that validates 301 new exoplane*ts, in «The Astrophysical Journal», 926-2 (2022), p. 120.

3 - N.J. Szymanski., Rendy, B., Fei, Y. et al., *An autonomous laboratory for the accelerated synthesis of novel materials*, in «Nature» 624 (2023), pp. 86-91 (https://doi.org/10.1038/s41586-023-06734-w).

4 - L. Messeri - M.J. Crockett, *Artificial intelligence and illusions of understanding in scientific research* in «Nature», 627-8002 (2024), pp. 49-58.

comprensione scientifica è reso ancor più pressante dal fatto che le tecniche di Machine Learning (ML), ovvero di apprendimento automatico, spesso operano come 'scatole nere' (black-box), la cui opacità sembra ostacolare il raggiungimento di una comprensione autentica[5].

La comprensione scientifica costituisce il fine ultimo della scienza, ben al di là della mera capacità predittiva. In questo contributo tratteremo la comprensione scientifica come una delle virtù epistemiche o intellettuali più elevate. La filosofia delle virtù intellettuali o epistemiche è un insieme eterogeneo di posizioni in cui la nozione di virtù intellettuale ricopre un ruolo centrale o nel risolvere grandi questioni epistemologiche tradizionali o nel tracciare nuove direzioni nella epistemologia, enfatizzando le abilità intellettuali nel soggetto epistemico[6]. Le virtù epistemiche più comunemente discusse sono l'accuratezza, la precisione, il rigore logico, la responsabilità intellettuale, il pensiero critico, e anche l'empatia verso i propri peer epistemici e il coraggio nel sostenere le proprie idee[7]. Grimm ha avuto il merito di interpretare la comprensione scientifica proprio come una virtù intellettuale[8]. La ragione per la quale in questo contributo si tratterà la comprensione scientifica come virtù intellettuale è che il suo interesse non si limita esclusivamente all'ambito scientifico, ma si estende a una riflessione più ampia sul valore dell'IA per la comprensione del mondo, del quale la comprensione scientifica è sicuramente un aspetto cruciale, e quindi che, se analizzato, può essere informativo sugli altri ambiti in cui la comprensione è cruciale. Adottare una concezione della comprensione così ampia come virtù intellettuale, dunque, ci permette di discutere come l'IA non solo sia strumentale per la ricerca di una verità scientifica o una comprensione scientifica, ma possa anche essere generatrice di virtù che trascendono quelle meramente proprie dell'ambito scientifico.

In particolare, nel presente contributo, si intende affrontare questa problematica in tre passaggi distinti. In primo luogo, verrà inquadrata la concezione di comprensione scientifica attraverso il prisma delle virtù intellettuali, evidenziando come queste possano offrire una chiave di lettura 'pienamente umana' per valutare il ruolo dell'IA nella pratica scientifica. Infine, si esaminerà se e in che misura le

---

5 - E. Sullivan, *Understanding from machine learning models* in «The British Journal for the Philosophy of Science» (2022) pubblicazione online first; F.J. Boge, *Two dimensions of opacity and the deep learning predicament*, in «Minds and Machines», 32 (1/2022), pp. 43-75; C. Beisbart - T. Räz, *Philosophy of science at sea: Clarifying the interpretability of machine learning* in «Philosophy Compass», 17 (2022/6), e12830.

6 - M. Croce, *Epistemologia delle virtù*, in «Aphex» 15 (2017).

7 - J. Baehr, *Educating for intellectual virtues: From theory to practice* in «Education and the growth of knowledge: Perspectives from social and virtue epistemology», (2013), pp. 106-123.

8 - S.R. Grimm, *Understanding as an intellectual virtue* in H. Battaly (ed.), *The Routledge handbook of virtue epistemology*, Routledge, New York 2018, pp. 340-351.

tecniche di Machine Learning possano contribuire alla comprensione scientifica. L'auspicio è che la prospettiva offerta in questo lavoro non si limiti all'ambito scientifico, ma offra spunti di riflessione anche in altri domini, dove l'IA sta assumendo un ruolo sempre crescente, e sempre più meritevole di attenzione filosofica.

Prima di addentrarci nella nostra analisi, è opportuno chiarire alcuni aspetti fondamentali. Innanzitutto, sebbene esistano molteplici forme di intelligenza artificiale, il presente contributo si concentrerà sul Machine Learning, ovvero alle tecniche di apprendimento automatico[9]. A tal proposito, la nostra analisi non discuterà se il machine learning può raggiungere, in quanto *artificial mind*, una reale comprensione scientifica, bensì se gli scienziati, in quanto soggetti epistemici, utilizzando tali strumenti di apprendimento automatico, possano raggiungere comprensione scientifica nonostante la loro opacità. In secondo luogo, è opportuno sottolineare che esistono varie tipologie di opacità che caratterizzano questi sistemi. La cosiddetta black box o opacità del machine learning si riferisce al fatto che, pur conoscendo gli input and osservando gli output, non sia possibile interpretare il processo che ha condotto a tali risultati. Questa opacità però può derivare o dalla difficoltà degli agenti epistemici nel comprendere i meccanismi interni del modello, o all'impossibilità di capire la funzione 'target' che il modello computa[10]. Queste distinzioni non saranno discusse ulteriormente perché, è bene ribadirlo, la questione di questo contributo non è principalmente sulla nostra comprensione dell'IA, bensì sulla possibilità che l'IA possa fornirci una comprensione del mondo, nonostante la sua parziale natura opaca e non interpretabile. Mentre la prima domanda necessiterebbe di un'analisi approfondita dei vari tipi di opacità coinvolti, il nostro quesito, pur intrecciandosi col primo, è parzialmente indipendente, riguardando, appunto, la comprensione del mondo nonostante l'opacità dell'IA. Infine, il nostro quesito può essere formulato in due modi distinti. Nel primo modo, il quesito posto è se l'IA si limiti ad aiutarci nella comprensione del mondo, fungendo da mero strumento facilitatore per il raggiungimento di tale obiettivo. Nella seconda accezione, invece, viene chiesto se l'IA fornisca una comprensione del mondo in senso proprio. In questo contributo terremo in considerazione entrambi i tipi di quesiti.

9 - In particolare, si discuteranno due esempi di Machine Learning, ovvero il GPTune, basato sull'utilizzo di modelli surrogati e ottimizzazione Bayesiana, e l'XGBoost, basato su alberi decisionali potenziati dalla tecnica del *gradient boosting*. Entrambi gli algoritmi, dunque, non si avvalgono di reti neurali.

10 - T. Räz, *ML interpretability: Simple isn't easy*, in «Studies in history and philosophy of science», 103 (2024), pp. 159-167.

La comprensione scientifica come virtù intellettuale

È senza dubbio un'opinione ampiamente condivisa che uno dei fini ultimi della scienza sia quello di offrire una comprensione scientifica del mondo. Indiscutibile è anche il fatto che, fino ad oggi, tale obiettivo sia stato ampiamente raggiunto. Grazie alla fisica, comprendiamo i processi di reazione nucleari; grazie alla genetica possiamo capire perché alcune patologie vengono tramandate da generazione in generazione; grazie alla geologia, invece, conosciamo le ragioni per cui alcune zone del pianeta Terra sono più suscettibili a terremoti o a eruzioni vulcaniche. Non si può dunque negare che la nostra comprensione del mondo sia stata profondamente ampliata dalla scienza.

Ma che cosa intendiamo esattamente con l'espressione 'comprensione scientifica'? Tale quesito ha ricoperto un ruolo centrale nel dibattito filosofico contemporaneo a partire dal celebre contributo di Michael Friedman "Explanation and Scientific Understanding" (1974), e dall'importante saggio di Salmon "What is Scientific Understanding?". Tale rilevanza stride con la scarsissima attenzione che era stata dedicata alla comprensione scientifica nel periodo precedente. Per Hempel (1965), la comprensione scientifica appariva come una nozione dal carattere pragmatico, soggettivo e psicologico, e dunque irrilevante o fuorviante per la scienza. La scienza, invece, si fonda su connessioni oggettive tra evidenza e teorie, e deve dunque trattare nozioni più oggettive come la spiegazione scientifica, basata, per Hempel, su nessi inferenziali tra le leggi di natura e fenomeni particolari[11]. In questa prospettiva, si riteneva che non vi fosse dunque alcuna dimensione della comprensione scientifica meritevole di attenzione, che non fosse già compresa nel concetto di spiegazione scientifica: fornire spiegazioni equivaleva a garantire comprensione scientifica. Tale prospettiva riduzionista ha trovato eco e sostegno anche in autori più recenti come Trout, il quale ha nuovamente sminuito la comprensione scientifica, interpretandola come un mero sentimento soggettivo, un momento di 'eureka', o una sensazione di 'soddisfazione intellettuale' provata dagli scienziati di fronte a una spiegazione[12].

Sebbene il dibattito sulla comprensione scientifica ricopra ora un ruolo prominente, e senz'altro vi sono filosofi che ammettono la natura poliedrica e contestua-

11 - C. Hempel, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, Free Press, New York 1965; C. Hempel, *Philosophy of Natural Science*, Englewood Cliffs, Prentice Hall New York 1966.

12 - J.D. Trout, *Scientific Explanation And The Sense Of Understanding*, in «Philosophy of Science», 69 (2/2002), pp. 212-233 (doi:10.1086/341050).

le della comprensione scientifica[13], la nozione di comprensione scientifica che viene discussa è prevalentemente di tipo esplicativo, ovvero fondata sulla capacità di rispondere a domande sul 'perché' di determinati fenomeni. In particolare, secondo la formulazione di comprensione scientifica articolata nel contributo di de Regt, la nozione di comprensione scientifica si articola in due piani diversi ma strettamente legati, ed entrambi di tipo esplicativo[14]. Da una parte, vi è la comprensione esplicativa di un fenomeno, dall'altra parte, vi è la comprensione esplicativa di una teoria. Una comprensione esplicativa di un fenomeno, infatti, richiede la comprensione di una teoria, e affinché una teoria possa essere compresa, essa deve essere intellegibile. Il concetto chiave della comprensione scientifica diventa dunque quello di intelligibilità delle teorie scientifiche. Senza intelligibilità delle teorie scientifiche non vi può essere comprensione scientifica.

Questo modo di approcciare la comprensione scientifica è certamente meritevole perché enfatizza il fatto che si stia parlando di comprensione scientifica, basata, tipicamente, su teorie scientifiche costituite da leggi della natura che rilevano la vera descrizione del mondo fenomenico. Il rischio però è che tale concetto di comprensione scientifica sia riduttivo. Si pensi a campi o discipline scientifiche dove ancora manca un impianto teorico valido e corroborato, come la fisica delle altissime energie, dove abbiamo solo modelli approssimativi, o in campo climatico o economico. In tali ambiti scientifici, sebbene gli scienziati lavorino con modelli che possono svolgere un ruolo teorico, in realtà essi non costituiscono un impianto epistemico solido come le teorie scientifiche vere e proprie. Sono piuttosto modelli ipotetici e provvisori. Inoltre, tale concezione di comprensione scientifica è limitante perché pone l'enfasi sulla teoria stessa, che deve essere intellegibile, piuttosto che sull'agente epistemico stesso. Ovviamente il vantaggio di tale definizione è di fondarsi su un aspetto 'oggettivo' di teorie scientifiche e dunque di evitare elegantemente ogni possibile critica riguardo il carattere soggettivo e psicologico della comprensione scientifica. Ma il rischio viene annullato, a mio avviso, con una manovra infelice, che sopprime la dimensione più virtuosa del soggetto epistemico che comprende. Qual è la condizione epistemica del soggetto che comprende e il valore epistemico della comprensione? Il mio obiettivo, in questo contributo, è di distogliere lo sguardo da una concezione di comprensione scientifica come il frutto dell'applicazione di una teoria intellegibile a un caso particolare, e di invece interpretarla come una manifestazione di una virtù propria del pensatore-scienziato, che indipendentemente da una teoria scientifica, riesce a comprendere il fenomeno indagato.

13 - H.W. de Regt - D. Dieks, *A Contextual Approach to Scientific Understanding*, in «Synthese», 144 (1/2005), pp. 137-170 (doi: 10.1007/s11229-005-5000-4).

14 - H.W. de Regt, *Understanding scientific understanding*, Oxford University Press, Oxford 2017.

Questo 'atto epistemico', tra l'altro, potrebbe essere cruciale e prioritario rispetto alla creazione di una teoria scientifica. In particolare, in questo contributo tratteremo la comprensione scientifica come una virtù intellettuale/epistemica.

Fin dalla pubblicazione di "Virtues of the Mind" di Zagzebski (1996), l'epistemologia delle virtù ha ampliato gli orizzonti dell'epistemologia, spostando l'attenzione dal valore epistemico della conoscenza ai cosiddetti beni epistemici superiori, quali la comprensione e la saggezza[15]. In questo senso, dunque, l'epistemologia abbraccia l'intera gamma delle 'eccellenze intellettuali' che ci permettono di cogliere le strutture della realtà e sviluppare un sapere saggio. Il soggetto che comprende il mondo (anche) a livello scientifico manifesta senza dubbio una eccellenza intellettuale. Se dunque l'IA fornisce o contribuisce a questa comprensione, il suo ruolo può essere visto come un potenziamento di tale virtù epistemica dell'uomo. Questo è un primo punto da indagare: fino a che punto l'IA può rafforzare tale virtù? Come abbiamo visto nella introduzione, una delle preoccupazioni principali è che l'IA eccella nel suo carattere predittivo, ma non necessariamente nel suo ruolo di fornire una comprensione scientifica. In ambiti in cui l'IA non si dimostra neppure particolarmente abile nelle predizioni scientifiche, come per esempio in ambito finanziario, si tende infatti a privilegiare un'analisi 'umana' e manuale dei dati, mentre l'uso dell'IA viene proprio abbandonato. Ciò è segno del fatto che l'IA non abbia potere di comprensione. In secondo luogo, se adottiamo la prospettiva del 'responsabilismo epistemico', la vera virtù intellettuale implica 'un'agenzia' (agency) attiva da parte del soggetto che quindi è responsabile della propria comprensione[16]. L'agente virtuoso non è colui che riceve passivamente la comprensione, ma colui il quale partecipa attivamente alla sua acquisizione. Sorge dunque un altro interrogativo: che ruolo svolge l'IA nella comprensione scientifica, sempre che svolga un ruolo? Rischia di sostituire l'agenzia intellettuale dell'uomo? Questa inquietudine si riflette anche nella vita quotidiana: si pensi, ad esempio, al bambino che svolge i compiti con l'ausilio di un algoritmo di machine learning. Il timore è che, affidandosi a uno strumento esterno, egli non sviluppi appieno la propria virtù intellettuale del comprendere. Infine, trattare la comprensione scientifica come una virtù intellettuale consente di oltrepassare i confini della filosofia della scienza e di esplorare il ruolo dell'IA in altri domini della conoscenza.

A questo punto, è necessario chiarire come possiamo concettualizzare la comprensione scientifica in termini di virtù epistemica.

---

15 - L.T. Zagzebski, *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*, Cambridge University Press, New York 1999 (doi:10.1017/CBO9781139174763).

16 - Per una discussione sulla differenza tra responsabilismo e affidabilismo epistemico, si veda L. Code, *Epistemic Responsibility*, University Press of New England, Hanover, NH 1987.

Secondo Grimm, la comprensione è la capacità di cogliere le connessioni tra i fenomeni indagati. Sebbene non esista una definizione di comprensione universalmente accettata in letteratura, essa è tradizionalmente intesa come un atto di cogliere o afferrare le relazioni strutturali, modali e di dipendenza che intercorrono tra gli elementi chiave di un dato dominio[17]. Ad esempio, per comprendere il cambiamento climatico, è necessario afferrare le relazioni tra effetto serra, emissioni di gas serra, aumento delle temperature, eventi meteorologici estremi, acidificazione degli oceani e altri fattori. Analogamente, comprendere un fenomeno come il fuoco significa cogliere la relazione tra combustibile, ossigeno e calore.

Un elemento centrale nella concezione della comprensione di Grimm dunque è il concetto di relazioni di dipendenza, che collegano variabili e sono suscettibili di cambiamento. Poiché l'oggetto della comprensione può essere rappresentato come una mappa causale o modale, una componente chiave del comprendere consiste nel rappresentare accuratamente queste relazioni di dipendenze tra variabili diverse. Queste mappe sono dinamiche piuttosto che statiche: le variabili possono assumere diversi valori, ciascuno dei quali influisce sul sistema in modo differente. Chi possiede una comprensione genuina è in grado di anticipare e prevedere il comportamento del sistema al variare di tali valori. Comprendere significa dunque afferrare o cogliere lo spazio modale del sistema e rispondere a interrogativi controfattuali, come "E se le cosse fossero andate diversamente?"[18]. Ciò implica anche la capacità di discernere quali variabili siano realmente rilevanti e quali invece irrilevanti nel contesto di analisi. Un individuo dotato di comprensione è quindi in grado di prevedere, inferire conseguenze modali e persino esercitare un certo controllo sul sistema, cogliendo, nella sua unitarietà, coerenza strutturale e interconnessione. Un'altra caratteristica della comprensione in quando virtù intellettuale è quella di essere un concetto 'graduale', in quanto ammette di essere 'posseduto' in gradi diversi. Questo ovviamente consente di argomentare che anche i bambini o gli animali possono acquisire gradi di comprensione del mondo, seppure non siano capaci di afferrare tutta la struttura di mappa modale.

La comprensione, dunque, si fonda su due elementi centrali. Da una parte, vi è il suo oggetto, che consiste in una mappa di relazioni modali che strutturano un

---

17 - Per una discussione puntuale della concezione di comprensione come virtù intellettuale si veda: S.R. Grimm, (2017). *Understanding and transparency*, in C. Baumberger - S.R. Grimm - S. Ammon (eds.), *Explaining understanding. New perspectives from epistemology and philosophy of science*, Routledge, London 2017, pp. 212-229. S.R. Grimm, *Understanding as an intellectual virtue*, cit., pp. 340-351; S.R. Grimm, *Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology*, Oxford University Press, New York 2019 (doi:10.1093/oso/9780190860974.001.0001).

18 - J. Woodward, *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press, New York 2003. (doi:10.1093/0195155270.001.0001).

sistema coerente e coeso, le cui varie componenti, dunque, sono integrate tramite nessi di dipendenza. Dall'altra parte, vi è l'atto epistemico, e quindi, inevitabilmente cognitivo, mediante il quale il soggetto afferra tale struttura relazionale. La comprensione è, in questo senso, fattiva, poiché presuppone che le relazioni modali che vengono colte rappresentino effettivamente, anche se magari solo approssimativamente, il mondo in modo veritiero. Al contempo, la compresione comporta un aspetto psicologico, dato che richiede che l'agente cognitivo colga i nessi modali che strutturano il fenomeno.

Il Machine Learning e la comprensione scientifica

L'impiego degli algoritmi di apprendimento automatico, detto anche machine learning sta diventando sempre più pervasivo, assumendo funzioni di primaria importanza, tra i quali spiccano i suoi ruoli di ottimizzazione sperimentale, predizione e analisi dati. Piuttosto che soffermarsi su un'analisi tecnica della loro natura e del loro funzionamento, il presente contributo offrirà un esempio concreto della loro applicazione, soffermandosi sul loro ruolo epistemico con una particolare attenzione a ciò che questi algoritmi offrono in termine di comprensione scientifica. Come detto precedentemente, questa scelta è motivata dal fatto che l'obiettivo della nostra indagine è interrogarsi non tanto sulla nostra capacità di comprendere tali algoritmi, bensì sulla misura in cui essi stessi contribuiscano alla comprensione scientifica. In particolare, il presente lavoro esaminerà l'utilizzo da parte dei fisici del Brookhaven National Laboratory (BNL) di due algoritmi specifici di machine learning, da un lato il GPTune per ottimizzare i risultati sperimentali, e dall'altro XGBoost per analizzare e interpretare i dati ottenuti dal processo di ottimizzazione. Questi due algoritmi possono, in qualche modo, fornire o quanto meno facilitare la comprensione scientifica del mondo?

Il caso specifico che discuteremo è l'uso di questi due algoritmi per massimizzare l'intensità del fascio ionico durante le fasi di iniezione e di estrazione dalla sorgente EBIS, una sorgente che produce ioni altamente carichi sfruttando il principio della ionizzazione per impatto elettronico, ovvero attraverso la collisione di elettroni con ioni[19]. Questo processo di produzione di ioni altamente carichi richiede delle condizioni sperimentali molto precise, che coinvolgono numerosi parametri. La sfida degli scienziati, dunque, è quella di trovare la combinazione ottimale di parametri, regolandoli manualmente. Essendo il numero di parametri molto elevato, e dato che i parametri presentano interazioni molto complesse tra di loro, il pro-

19 - G. Zschornacka - M. Schmidt - A. Thorn, *Electron Beam Ion Sources,* 2014, in «*arXiv*» (arXiv:1410.8014).

cesso è estremamente complicato e dispendioso di energie e di tempo. È proprio in questa fase, dunque, che l'apprendimento automatico fornisce un valore inestimabile. In particolare, il GPTune, che è un framework di autotuning basato sull'ottimizzazione di tipo bayesiana, è estremamente efficiente per identificare i valori ottimali dei parametri e ottenere una ottimizzazione della configurazione complessiva di essi[20]. Esso cerca in modo intelligente la combinazione ottimale di parametri esplorando il loro spazio configurazionale, identificando la regione dello spazio configurazionale che presenta maggiori probabilità di contenere la migliore combinazione di parametri per l'ottimizzazione del fascio, e regolando iterativamente i parametri del sistema in modo autonomo. Questo processo si sviluppa e in tre fasi diverse. Nella prima fase, detta di campionamento, una volta impostato l'obiettivo di ottimizzazione del fascio, il GPTune campiona casualmente alcune configurazioni iniziali di parametri e li valuta; questi campioni 'pilota' forniscono una prima conoscenza di come i parametri influenzino l'obiettivo. Nella seconda fase, detta di modellizzazione, il GPTune utilizza la valutazione iniziale della fase di campionamento per addestrare il un modello di probabilità per la cosiddetta funzione obiettivo in modo non supervisionato. In terzo luogo, il GPTune aggiorna le sue probabilità e propone nuove configurazioni che hanno maggiori probabilità di ottimizzare il fascio[21]. Questo processo avviene in tempo reale ed è completamente automatizzato. È un meccanismo di feedback, nel quale l'apprendimento avviene in modo iterativo, funzionando con la raccolta continua di feedback, aggiornandosi, regolando e modificando i parametri per migliorare le prestazioni del sistema. Il GPTune si è rivelato uno strumento di grande successo per la ottimizzazione dell'intensità del fascio sia per la fase di iniezione sia per quella di estrazione del fascio, raggiungendo addirittura un notevole incremento dell'intensità del fascio pari al 70% in un certo punto di misurazione.

Sebbene il GPTune sia uno strumento altamente efficace nell'ottimizzazione in tempo reale della intensità del fascio ionico, esso da solo non è sufficiente a comprendere e interpretare i dati prodotti dal processo di ottimizzazione. GPTune si concentra solo sulla ricerca della configurazione di parametri ottimale, ma non fornisce una analisi delle correlazioni tra le variabili sperimentali né una spiegazione dettagliata dei fenomeni fisici. Per poter capire il meccanismo di ottimizzazione e comprendere il sistema fisico, i fisici del Brookhaven National Laboratory affiancano al GPTune un altro potente algoritmo di apprendimento automatico, che è l'XGBoost. Questo modello, diventato celebre per essere stato il vincitore dell'Higgs Boson Machine Learning Challenge nel 2014, si basa sulla tecnica dell'*extreme*

20 - Cho et al., *GPTune user guide,* 2022, in gptune.lbl.gov/documentation/gptune-user-guide/

21 - *Ibidem.*

gradient boosting e costruisce modelli predittivi basati su alberi di decisione. Esso funge da modello black-box, e svolge il ruolo di analizzare la grande quantità di dati sperimentali raccolta da GPTune, interpretarli individuandone pattern nascosti e relazioni complesse tra i parametri e l'intensità del fascio, e infine predire l'intensità del fascio in base ai parametri sperimentali, permettendo dunque di valutare in anticipo gli effetti di possibili regolazioni dei parametri. Infatti, addestrandosi sui dati ricevuti dal GPTune, l'XGBoost è in grado di predire con una accuratezza del 79% per l'iniezione e dell'80% per l'estrazione, mostrando dunque una piena convergenza tra il modello e il sistema fisico[22]. XGBoost però rimane un modello black-box, non riuscendo a fornire una spiegazione dei processi fisici sottostanti. Sebbene sia utile per individuare correlazioni significative tra i parametri e l'intensità del fascio, nonché pattern dei dati sperimentali, non può sopperire a una mancanza di comprensione scientifica del sistema fisico. I fisici, dunque, non dispongono di una conoscenza esplicita e dettagliata del perché e di come determinate combinazioni di parametri conducano a una ottimizzazione del fascio rispetto ad altre. Rimane loro ignota la relazione specifica tra tali configurazioni e l'intensità del fascio, così come manca una spiegazione trasparente dei processi fisici che si verificano durante le fasi di iniezione e estrazione da EBIS. Ne consegue che non possono neppure sfruttare l'analisi di dati per formulare autonomamente predizioni future, senza l'ausilio di XGBoost. Per ottenere un livello di comprensione maggiore, seppur parziale, è imprescindibile l'utilizzo di strumenti di *explainable-AI*, che sono capaci di caratterizzare e quantificare l'influenza di ciascuno parametro e quella della configurazione totale di parametri sulla intensità del fascio. Dunque, per interpretare le previsioni dell'XGBoost, i fisici lo integrano con strumenti come i grafici di dipendenza parziale, i *partial independence plot* (PDP), e le Shapley Additive exPlanations, (SHAP).

I grafici di dipendenza parziale (PDP) visualizzano gli effetti marginali di uno o due parametri di input sul risultato previsto. Aiutano a illustrare in che modo un singolo parametro sperimentale influisce sull'output stimato mantenendo costanti gli altri parametri. I valori SHAP *quantificano* il contributo marginale di ciascun parametro di input alla previsione del modello in modo più robusto, in quanto prendono in considerazione il fatto che i diversi parametri possono essere correlati. SHAP fornisce informazioni preziose sulle relazioni tra i diversi parametri modellati da XGBoost in quanto identifica l'importanza relativa di varie caratteristiche correlate.

22 - Gu et al., *Enhancing beam intensity in RHIC EBIS beamline via GPTune machine learning-driven optimization,* No. BNL-226095-2024-COPA, Brookhaven National Laboratory (BNL), Upton, NY - United States, 2024.

Conclusioni

Dall'analisi fin qui svolta, emergono considerazioni di notevole rilevanza per l'oggetto della nostra indagine, ovvero il contributo del machine learning per la comprensione scientifica intesa come virtù intellettuale. Il machine learning svolge un ruolo sperimentale straordinario per l'ottimizzazione delle condizioni sperimentali, ma anche per l'analisi dei dati e la formulazione di predizioni. Tuttavia, la capacità di analisi, l'interpretazione di dati, così come quella predittiva rimangono preliminari e insufficienti rispetto al raggiungimento di una comprensione scientifica autentica e piena del sistema scientifico. Se però ammettiamo, come lo abbiamo fatto, che la comprensione scientifica possa manifestarsi per gradi, lo scenario non è completamente negativo.

Innanzitutto, dal caso esaminato, si può evincere che la combinazione GPTune + XGBoost + *explainable AI* contribuisca indubbiamente a una acquisizione abbastanza profonda della struttura modale del fenomeno fisico sotto investigazione, ovvero quello del fascio ionico. In particolare, gli strumenti di *explainable AI* come SHAP o i PDP sono imprescindibili per poter afferrare le relazioni modali tra le configurazioni dei parametri e l'intensità del fascio, così come tra i parametri singoli e l'intensità del fascio. Le informazioni che gli scienziati ricavano sono molto importanti dal punto di vista modale, perché rivelano quali parametri, con quali valori e in quale misura influenzano maggiormente il fascio e contribuiscano alla sua ottimizzazione. Tali strumenti dunque, attenuano l'opacità del black box del XGBoost e rivelano connessioni modali altrimenti inaccessibili. Essendo tali relazioni modali l'oggetto proprio della comprensione scientifica, è dunque possibile concludere che l'uso degli *explainable AI* fornisce un certo grado di comprensione scientifica. Ovviamente è bene precisare che neppure gli strumenti di *explainable AI* forniscono una mappa modale esaustiva, che è ciò che costituirebbe una comprensione scientifica piena. Due sono i motivi principali di questa limitazione. In primo luogo, la selezione dei parametri per il funzionamento di GPTune e l'analisi di XGBoost è limitata, essendo il numero di parametri coinvolti nel processo fisico di ionizzazione estremamente elevato. Secondariamente, strumenti come SHAP o PDP non offrono indicazioni definitive sulla natura modale delle relazioni di dipendenza. Tuttavia, l'*explainable AI* garantisce una comprensione parziale, consentendo la costruzione – seppur incompleta e provvisoria – di relazioni modali.

A questo punto, possiamo porci un ulteriore interrogativo: se il machine learning rimanesse totalmente inintelligibile e non-interpretabile, potremmo raggiungere comunque almeno un livello minimo di comprensione scientifica? Seguendo il modello di de Regt, la comprensione scientifica è possibile solo in presenza di una teoria intellegibile. Se dunque applicassimo tale principio di intelligibilità all'IA, dovremmo concludere che solo un AI trasparente può realmente fornire

comprensione scientifica. Ciò non significherebbe che un sistema di machine learning opaco sia privo di valore euristico, ma comunque non fornirebbe comprensione scientifica. In questa ottica, l'assenza di strumenti esplicativi impedirebbe infatti di chiarire le relazioni dinamiche sottostanti, limitandoci a individuare mere correlazioni empiriche tra input e output, prive di una interpretazione teorica anche solo minima. L'IA in questo caso procura una ottimizzazione dei parametri, fornendo indicazioni su quali combinazioni di parametri risultino più efficaci per raggiungere un'alta intensità del fascio ionico, ma senza fornire una vera e propria intelligibilità del processo fisico. Tuttavia, forse, una tale conclusione è troppo pessimista, e nasconde il fatto che le predizioni stesse condotte dall' XGBoost consentono di delineare delle relazioni (empiriche) modali tra configurazioni di parametri e intensità del fascio. Si potrebbe dunque argomentare che anche l'XGBoost stesso possa fornire una intuizione o conoscenza debole di alcune relazioni modali, e quindi un primo livello di comprensione scientifica. Oppure, forse, più solidamente, si potrebbe argomentare che, seppure non fornendo comprensione scientifica, anche un modello di machine learning black-box possa *aiutare* o *facilitare* il raggiungimento di comprensione scientifica.

Date queste conclusioni ottimiste sul ruolo del machine learning nella scienza, in quanto il machine learning riuscirebbe, se interpretabile, a fornire una buona comprensione scientifica, e, se non interpretabile, comunque a facilitare tale processo di comprensione scientifica, occorre comunque ricordare un importante limite. Le condizioni sperimentali che si ottengono in un laboratorio di EBIS, e in un qualunque laboratorio di fisica delle particelle, sono spesso estremamente 'fragili' e variabili a seconda del contesto specifico nel quale viene effettuato l'esperimento. Ciò implica che le informazioni estratte da un sistema potrebbero non essere replicabili in un sistema analogo, costruito in un momento successivo, o in un luogo diverso. Pertanto, le relazioni modali che si potrebbero ricavare dall'utilizzo di strumenti di IA potrebbero non essere 'robuste', ovvero potrebbero non rivelarsi veritiere in altri contesti. Questo però, è bene enfatizzare, non riguarda un problema dell'IA, quanto piuttosto la non-replicabilità dei sistemi fisici creati in laboratorio.

Infine, si pone la questione della responsabilità epistemica dello scienziato e della scienziata nella costruzione della propria comprensione scientifica. Se l'IA fornisce risultati autonomamente, senza richiedere una interazione attiva da parte degli scienziati, possiamo ancora parlare di un processo di comprensione genuina? Confrontiamo questo scenario con quello che si avrebbe senza un uso dell'IA. In quest'ultimo caso, gli scienziati e le scienziate dovrebbero manualmente cercare di cambiare le configurazioni dei parametri sperimentali fino a trovare quella ottimale. In questo processo, questi si sforzerebbero di capire delle relazioni modali tra i parametri che regolano manualmente e l'intensità del fascio. Sarebbero dunque

pienamente attivi a livello non solo pratico ma anche epistemico. L'IA, invece, offre informazioni senza che il soggetto epistemico intervenga direttamente nella loro elaborazione. Si tratta di un processo largamente, se non completamente, automatizzato di cui il ricercatore-scienziato sembra usufruire a livello passivo. A questo punto, però, la risposta parzialmente pessimista data precedentemente, ovvero che l'IA o ci fornisce solo una comprensione molto parziale, o semplicemente ci faciliti il suo raggiungimento, diventa un aspetto positivo nella questione sulla passività degli scienziati e delle scienziate nell'uso dell'IA. Se davvero aspiriamo a una scienza che fornisca comprensione scientifica piena, l'IA può solo essere concepita come un primo passo per la costruzione di comprensione scientifica, in quanto non può sostituire in alcun modo il processo di comprensione scientifica. L'IA dunque non esime e non solleva lo scienziato e la scienziata dal compito di comprendere attivamente. Se da un lato questa constatazione potrebbe apparire poco rassicurante perché implica il rischio che, qualora ci si accontenti di solo IA, la scienza diventerebbe un mero strumento predittivo, dall'altro è consolante, perché ribadisce che l'IA non annulla né sovverte l'*agency* epistemica umana. Se la scienza vuole rimanere la disciplina che fornisce una comprensione scientifica del mondo, allora preserverà il suo carattere intrinsecamente umano.

Adottare un concetto di comprensione scientifica come virtù intellettuale, intesa come capacità di afferrare relazioni modali in un sistema, si rivela, a mio avviso, fecondo, in quanto permette di applicare le conclusioni di questo articolo in altri ambiti in cui l'IA potrebbe essere utilizzato per contribuire o facilitare il processo di comprensione del mondo. L'auspicio è dunque che le tesi offerte come frutto di questa riflessione aprano la strada a considerazioni sul ruolo epistemico che l'IA svolge nella crescita delle virtù intellettuali dell'agente epistemico, in tutti i campi nei quali queste operano.

VERA MATARESE
Ricercatrice a tempo determinato in Logica e filosofia della scienza presso il Dipartimento di Filosofia, Scienze sociali, Umane e della Formazione dell'Università degli studi di Perugia
vera.matarese@unipg.it

L'innovazione tecnologica dell'IA nell'età della tecnica

di Sergio Rondinara

The technological innovation of AI will be able to help the human person in various aspects of his personal and social life. But it will also cause problems already intrinsic to the innovation and that will be amplified by the marked cultural context of the domain of technology. This essay proposes the search for a line of thought to support an ethically responsible strategy for the advent of technological innovations.

Keywords: Artificial Intelligence; Technological Innovation; Technique; Wisdom; Speculative Rationality

Sono innumerevoli i caratteri culturali che contraddistinguono la complessità dell'epoca a noi contemporanea. Risulta pertanto estremamente difficile, se non impossibile, farne un'analisi che la esprima pienamente o la spieghi in modo esaustivo. Non possiamo però non notare l'ambiguità e il disorientamento provocato, da un lato, dallo stato d'animo di sentirci al vertice di un processo di civilizzazione avviato da alcuni millenni, e dall'altro nell'esser immersi in quel particolare processo di strutturazione delle società umane e dell'ambiente naturale raggiunto attraverso lo sviluppo tecnologico, che per la sua incisività non ha riscontro in epoche passate. Ci sentiamo fortemente legati a tutti quei processi culturali che partendo dagli albori delle nostre civiltà ci raggiungono e coinvolgono quotidianamente il nostro pensare tenendo così lo sguardo del nostro spirito rivolto all'indietro, ma allo stesso tempo, ci sentiamo sempre più padroni di noi stessi, della nostra storia e soprattutto del nostro futuro mediante il dominio tecnico che abbiamo imparato ad esercitare sul mondo. Siamo collocati in una spaccatura culturale in cui siamo, per così dire, rivolti all'indietro verso un passato in cui la saggezza (φρόνησις) era un valore ambito e pregnante, e nello stesso momento proiettati verso un futuro intrastorico programmabile e pianificabile attraverso la tecnica (τέχνη). Tutto ciò genera e segna, almeno in parte, quell'ambiguità e quel disorientamento che contraddistinguono la nostra epoca in cui viene a mancare un riferimento chiaro e univoco verso cui rivolgersi.

I concetti *tecnica* e *saggezza* formano oggi una diade che affiora sull'orizzonte filosofico con tutta la sua drammatica portata mostrando la grande rilevanza che ha per il futuro delle civiltà. Si tratta, dunque, del rapporto tra le nostre capacità di produrre mezzi che ci permettono di superare le innate carenze biologiche della specie umana e la nostra attitudine a esser capaci di riflettere profondamente sulle esperienze passate e presenti per poi esprimerne una valutazione esistenziale ed etica.

In tale contesto le stesse analisi delle società contemporanee rischiano di fare ancora ricorso a categorie che, inadatte a dire la rapidità dei cambiamenti, hanno condotto a definizioni tra loro diverse ma spesso con lo stesso prefisso "post": post-industriale, post-fordista, post-analogica, post-moderna.

Il progresso economico-sociale prodotto dall'industrializzazione mediante la realizzazione di innovazioni tecnologiche che con i loro prodotti stanno cambiando il mondo delle relazioni umane[1] ha prima favorito, e poi imposto, un dominio dello spirito tecnico-scientifico in quasi tutti i campi dell'agire umano. C'è da chiedersi se non stiamo facendo slittare il percorso dell'esperienza culturale della nostra

1 - Cfr. L. Floridi, *Infosfera. Etica e filosofia nell'età dell'informazione*, Giappichelli Editore, Torino 2009, pp. 185-194.

specie verso una progressiva razionalizzazione unidimensionale dello spirito umano determinata da uno sviluppo ipertrofico della razionalità strumentale e funzionale della tecnica penalizzando quella razionalità speculativa che presiede e regola le nostre domande sull'esistenza, sul senso e sui fini.

La questione che improrogabilmente si pone, è: fino a che punto vogliamo protrarre tale operazione di razionalizzazione che tende a spegnere in noi la capacità di porci le domande fondamentali sull'esistenza e ricercarne una soluzione autenticamente credibile? Saremo capaci, a partire da un domani prossimo, di una saggezza adeguata a tale sfida?

Dalla rivoluzione industriale ad oggi abbiamo nutrito una meraviglia tanto estatica quanto ingenua nei confronti dei prodotti della tecnica insieme con la fiducia di realizzare per mezzo di essa un mondo più umano. Ma dall'inizio del Novecento i dubbi hanno cominciato a crescere, hanno dapprima trovato voce nell'inquietante pensiero di qualche profeta e oggi sono diventate un coro di critiche soprattutto in ambito filosofico. Tutti segnali di un diffuso allarme culturale che segna il passaggio di una civiltà verso un suo nuovo assetto. Infatti, non stiamo forse varcando una soglia culturale in cui la nostra civiltà sta transitando da una civiltà e cultura industriale ad una civiltà e cultura post-industriale? È dall'avvio della rivoluzione industriale che ci troviamo all'interno di una tale transizione che dura ormai da circa due secoli e della quale non conosciamo ancora a fondo né le caratteristiche peculiari, né le potenzialità.

La nostra tradizione culturale mostra i segni inequivocabili di un lento volgersi verso il tramonto. In genere assistiamo ad un indebolimento della ragione e a un livellamento verso il basso dell'attività culturale. Non distinguiamo più con facilità un centro verso cui rivolgere lo sguardo, si ha come la sensazione di andare alla deriva. Siamo più inclini al rilassamento e alla fuga da quelle attività che postulano impegni seri o gravosi come l'inoltrarci in sentieri esigenti che conducono a modelli culturali carichi di sfide.

Viviamo un'epoca di transizione in cui insieme ai sintomi di una implacabile decadenza coesistono anche sintomi di una sanità reale come quelli di una critica che, con un profondo senso di preoccupazione, assiste culturalmente all'isolamento della persona umana e che tenta di dar voce a quegli spazi, come la sfera artistica o religiosa, che in una società segnata da logiche strumentali e funzionali non sono tanto considerati pericolosi o sospetti, quanto soprattutto inutili.

La nostra si caratterizza per essere un'epoca in cui i caratteri della civiltà che abbiamo alle spalle e della passata cultura s'intersecano e interferiscono con i motivi emergenti di una nuova cultura profondamente segnata dalla tecnica che non ha ancora ben abbozzato e delineato il suo discorso sull'umano.

La realtà della tecnica

Qual è la realtà della tecnica? La tecnica si costituisce e si esprime in tutte quelle attività mediante le quali vengono prodotti strumenti, mezzi e sistemi per migliorare le nostre condizioni di vita e di lavoro. Cioè quelle attività, capacità e mezzi con i quali la persona umana, nell'esplicare la propria esistenza, tenta di mettere i fenomeni naturali sotto il proprio controllo e la natura al suo servizio. La tecnica non è una primizia della nostra epoca, essa esiste da quando esistono le varie culture.

La tecnica è insita nella nostra natura, ci è "connaturale", in quanto il nostro agire, volto a migliorare e a produrre sempre nuove possibilità di sopravvivenza, è sempre stato sin dagli albori dell'umanità un "agire tecnico". La tecnica si caratterizza, pertanto, come uno dei più potenti fattori culturali che nel corso della storia ha permesso agli uomini e donne di ogni epoca di compensare le proprie carenze biologiche, di adattarsi ai vari climi e di colonizzare spazi a latitudini sempre maggiori. Dalla lavorazione della selce più rudimentale alla realizzazione dei più moderni e sofisticati sistemi informatici, la tecnica ci ha costantemente aiutato ad abitare il pianeta in condizioni di sempre maggiore sicurezza e confort, anche se non si può dimenticare che allo stesso tempo ha svolto un ruolo negativo nel contribuire all'allestimento di strumenti di aggressione e di morte.

Non possiamo dimenticare che il termine *tecnica* sta ad indicare non solo il mondo dei mezzi, o apparati, che nel loro insieme compongono il sistema tecnico, ma con esso si vuole soprattutto far qui riferimento a quella forma della razionalità umana che presiede al loro impiego in termini di ottimizzazione, concretezza pratica, funzionalità ed efficienza. Tutto ciò ci svela un aspetto molto importante della tecnica, il fatto che essa sia nata primariamente non come un'espressione creativa dello spirito umano, ma come il modo di trovare una soluzione alle nostre intrinseche carenze biologiche.

Nella lettura che stiamo facendo dell'epoca che viviamo non possiamo non riconoscere quanto la razionalità che presiede e governa la tecnica, la cosiddetta razionalità funzionale, tenda a vivere di vita propria, tenda ad affermare una radicale autonomia – in realtà è una vera e propria indipendenza –, tenda ad essere refrattaria ad ogni controllo etico, per plasmare e poi strutturare secondo i propri criteri e la propria scala di valori la mentalità di tutte quelle persone che entrano in contatto con essa.

Nell'età della tecnica, la razionalità che essa veicola ed afferma tende prima ad indebolire e poi ad abolire in noi quell'orizzonte umanistico all'interno del quale ci poniamo le domande di senso e significato sul mondo di cui siamo parte integrante, sulla nostra esistenza e sulla sfera del trascendente. In una parola la razionalità funzionale tende a ibernare, a devitalizzare la nostra *interiorità*, la nostra anima. In

tale contesto anche se riuscissimo a porci le domande di senso esse resterebbero inevase poiché la razionalità della tecnica non possiede in sé la capacità di dare risposte a tali domande e non perché essa non abbia ancora raggiunto nel suo sviluppo la piena maturità. La tecnica, infatti, è intrinsecamente incapace di condurci allo svelamento della verità, essa non tende a un fine, non dischiude scenari di senso. La tecnica ottimizza, calcola, fa.

La funzionalità e l'efficienza, che caratterizzano la razionalità della tecnica, sono diventate dei veri e propri riferimenti paradigmatici che segnano tutte le sfere dell'agire umano, sono diventate le direttrici principali e le forze operative che governano la vita dei popoli, fino al punto da attribuirsi il ruolo di un vero e proprio valore etico, e si propongono come guida e meta all'azione del singolo e della società.

L'efficienza in se stessa, affermando un intelligente e parco orientamento a un fine, non ha nulla di negativo, ma con il suo sviluppo la tecnica ha trasformato la stessa pratica dell'efficienza da un mezzo a un fine. È diventata un fine. Mediante questa inversione tra mezzi e fini la tecnica ha realizzato la sua metamorfosi. Da mezzo per superare le nostre carenze biologiche è diventata fine delle nostre azioni, da serva è diventata padrona rendendosi indipendente dalla persona umana.

Una tale inversione tra mezzi e fini non è connaturale alla tecnica, storicamente essa si è prodotta in epoca moderna. Anticipata e preparata dalla svolta cartesiana, in cui la ragione è diventata una ragione soggettiva – mentre da sempre si era essenzialmente configurata come una ragione oggettiva –, essa ha preso avvio dalla rivoluzione industriale in avanti. In epoca premoderna, in cui il reale era interpretato secondo una gerarchia di esseri, contavano inevitabilmente soprattutto i fini, gli scopi. Ma ora che nel nostro mondo contemporaneo si guarda soltanto ai mezzi, l'esito di tale processo è che il pensiero speculativo non serve più a sapere se un dato fine sia desiderabile in sé, perché questo non dipende più dalla ragione. La ragione, di fatto, è diventata strumento.

È questa la ben nota tesi sostenuta da Max Horkheimer ne *L'eclisse della ragione*: «La ragione è ormai completamente aggiogata al processo sociale; unico criterio è diventato il suo valore strumentale, la sua funzione di mezzo per dominare gli uomini e la natura»[2].

Con il decadere del discorso teleologico la ragione viene a perdere i suoi contenuti, diventa vuota, puro strumento di calcolo e il suo ruolo diventa esclusivamente formale. Per Horkheimer una tale ragione può essere definita solo come *stupidità* e la sua attività un vero e proprio processo di istupidimento.

La ragione funzionale della tecnica si sta progressivamente sostituendo in maniera egemonica alla ragione speculativa che presiede alla nostra riflessione sui

2 - M. Horkheimer, *L'eclisse della ragione*, Einaudi, Torino 1947, p. 25.

fini, sul senso del reale, sul significato dell'esistere; e una volta che l'avrà scalzata diventando predominante renderà la nostra ragione vuota, senza contenuto e mero strumento di calcolo. Fini e mezzi hanno invertito i ruoli, si sono dunque scambiate le parti: invece di produrre e ricercare i fini, lo scopo dell'uomo contemporaneo sembra paradossalmente essere diventato quello di produrre mezzi. E ciò che rende grave la situazione è che questi stessi mezzi tendono poi a giustificare i suoi fini.

Quale autonomia per la tecnica

Ma vi è un altro carattere preoccupante della tecnica: la sua autonomia. Un'autonomia che, come presto vedremo, si caratterizza essenzialmente come una vera e propria indipendenza. Un attento interprete della modernità contemporanea, Hans Jonas, ha così scritto a riguardo:

> […] gli sviluppi […] del fare tecnologico presentano la tendenza a rendersi autonomi, ossia ad acquisire una propria dinamica coattiva […] del quale non soltanto diventano irreversibili, […] ma acquistano una funzione propulsiva al punto da trascendere la volontà e i piani degli attori. […] ci sottrae di mano la legge dell'agire e […] mentre siamo liberi di fare il primo passo, al secondo e a tutti gli altri successivi siamo già schiavi[3].

Nel suo prorompente dinamismo carico di successi la tecnica con le sue innovazioni tecnologiche reclama una propria autonomia morale e pretende per sé un valore autonomo in quanto ritiene spurio e rifiuta qualunque riferimento assiologico o teleologico estrinseco al suo modo d'essere.

Ogni qual volta costatiamo all'interno della sfera culturale nella quale siamo immersi una progressiva sostituzione del giudizio qualitativo – tipico della nostra capacità di afferrare ed esprimere sia valori che significati e fini caratterizzanti la nostra esistenza – con un giudizio quantitativo tipico di un pensiero calcolante e della razionalità della tecnica, ci troviamo dinanzi a quell'atteggiamento autoreferenziale del sistema tecnico che sancisce la propria indipendenza da ogni sistema assiologico ad esso estrinseco.

Da questa sostituzione di giudizi nasce la disposizione dell'universo della tecnica ad escludere ogni sua dipendenza da un pensiero etico che non sia riconducibile alla propria forma di razionalità.

Inoltre, la tecnica offre all'uomo contemporaneo l'ideale di una umanità prometeicamente onnipotente, ma alquanto unidimensionale a motivo del suo appiattimento sulla sola razionalità funzionale. A causa di ciò i grandi quesiti sulla verità,

3 - H. Jonas, *Il principio responsabilità*, Einaudi, Torino 1979, p. 41.

sul senso dell'esistere, sul bene, perdono di significato, e il concetto di verità viene ad identificarsi con la pura e semplice nozione di *efficacia*. È vero ciò che funziona efficacemente.

Il sistema tecnico[4] esige per sé che gli venga riconosciuta non soltanto la capacità di poter formulare un proprio codice morale, e quindi una sua vera e propria legittimazione morale, ma soprattutto avanza la pretesa di una non negoziabile indipendenza rispetto ad ogni altro sistema etico o insieme valoriale emerso al di fuori del suo stesso ambito.

È proprio l'occultamento dei fini che la tecnica coltiva in sé a fare in modo che essa veicolando una razionalità unidimensionale, come quella funzionale, proponga di fatto una sua tipica irrazionalità.

Da questo quadro culturale sommariamente tratteggiato si può comprendere l'interrogativo allarmato di tanti filosofi del secolo scorso e a noi contemporanei[5]: dove potrebbe approdare l'umanità semmai ci fosse una resa senza condizioni del pensare umano alle pretese avanzate dalla tecnica? Una citazione per tutte:

> Ciò che è veramente inquietante non è che il mondo si trasformi in un completo dominio della tecnica. Di gran lunga più inquietante è che l'uomo non è affatto preparato a questo radicale mutamento del mondo. Di gran lunga più inquietante è che non siamo ancora capaci di raggiungere, attraverso un pensiero meditante, un confronto adeguato con ciò che sta realmente emergendo nella nostra epoca[6].

In questo scenario si colloca la grande sfida per l'uomo contemporaneo immerso nel sistema tecnico: il recupero della propria interiorità. Filosoficamente parlando, il recupero dell'anima mediante una cura conscia, intenzionale e particolareggiata della propria vita interiore. Cura che superando ogni insidia conformistica liberi la persona da quella *forma mentis* che inquadra ogni problema o questione in termini esclusivamente quantitativi, di ottimizzazioni ed efficienze varie. Il pensiero delle molteplici culture, la filosofia, l'arte nelle sue varie forme, la poesia, il sen-

---

4 - Cfr. J. Ellul, *Le système technicien*, Le Cherche-Midi Editeur, Paris 2004.

5 - Cfr. H. Arendt, *Vita activa. La condizione umana*, Bompiani, Milano 2008; J. Elull, *Il sistema tecnico*, Jaca Book, Milano 2009; U. Galimberti, *Psiche e techne. L'uomo nell'età della tecnica*, Feltrinelli Editore, Milano 2000[4]; A. Gehlen, *L'uomo nell'era della tecnica,* Armando Editore, Roma 2003; M. Heidegger, *La questione della tecnica*, in Id. *Saggi e discorsi*, Mursia, Milano 1991, pp. 5-27; M. Horkheimer, *Eclisse della ragione*, Einaudi, Torino 1970; H. Jonas, *Dalla fede antica all'uomo tecnologico*, Società editrice il Mulino, Bologna 1991; H. Jonas, *Il principio di responsabilità. Un'etica per la civiltà tecnologica*, Einaudi, Torino 1993; E. Severino, *Il destino della tecnica*, Rizzoli, Milano 1998; E. Severino, *Téchne. Le radici della violenza*, Rizzoli, Milano 2002; O. Spengler, *L'uomo e la tecnica. Contributo a una filosofia della vita*, Guanda, Parma 1992.

6 - M. Heidegger, *L'abbandono*, Il Melangolo, Genova, 1983, p. 36.

so della storia sono tutti ingredienti di quella formazione umanistica che, compensando il ruolo della tecnica nella vita culturale, contribuiscono sostanzialmente al recupero della nostra interiorità.

Non si tratta pertanto di rifiutare in blocco la tecnica per desiderare nostalgicamente un passato irripetibile, ma si tratta di recuperare quella vita interiore capace di produrre in noi quella saggezza che sappia coordinare l'uso cooperativo della razionalità speculativa e della razionalità funzionale, che sappia esercitare integralmente la nostra capacità contemplativa e la nostra capacità operativa e tecnica. La saggezza, dunque, come forma dell'intellettualità dell'uomo di domani che sappia valorizzare oltre ai valori umanistici trasmessi da una tradizione culturale anche quegli elementi antropologici (serietà del lavoro per il raggiungimento dei fini proposti, disincanto dalle troppo facili utopie sociali, educazione alla provvisorietà e alla rivedibilità) che la sfera della tecnica è capace di offrire alla vita interiore della persona umana.

Coltivando saggiamente la nostra interiorità potremmo tonificare, con l'esercizio, la coscienza di essere un centro autonomo e irripetibile di pensiero, volontà e libertà. In questo modo potremmo coltivare ciò che è più tipico della specie umana: ci scopriremo non solo capaci di affinare le nostre valutazioni ma di avere allo stesso tempo potenziato qualitativamente la nostra intrinseca predisposizione all'esercizio dell'alterità. La cura della vita interiore che nasce dalla saggezza e che a sua volta incrementa, Socrate ce lo ha testimoniato, è di per sé diffusiva e contagiosa. La cura di un'autentica interiorità inevitabilmente crea un fascino spirituale attraente, tende a produrre intorno a sé più consapevolezza, più attenzione e libertà.

È in quest'orizzonte di dominio della razionalità della tecnica che noi assistiamo all'apparire di innovazioni tecnologiche più o meno annunciate. L'Intelligenza Artificiale (IA) è quella che oggi domina la curiosità e l'informazione. Cerchiamo ora di capirne il programma senza entrare troppo nel dettaglio.

Che cos'è l'IA

Gli ultimi anni hanno visto una crescita costante dei programmi di ricerca nel campo dell'IA, diventata ormai un fenomeno di massa che suscita un interesse senza precedenti in ogni dove. Recentemente tutto ciò è stato ulteriormente rafforzato da prodotti informatici di enorme successo quali vari tipi di ChatGPT. Questi prodotti sono la concretizzazione visibile di decenni di ricerca nel campo delle reti neurali e della IA, con una accelerazione importante che dal 2012 le ha portate a diventare uno strumento di rilievo in tantissimi campi, dagli assistenti vocali all'analisi di proteine.

I programmi di ricerca riguardanti l'IA, nella loro essenza, sono da sempre un settore di studio e ricerca dell'informatica e questo sin dai suoi inizi con i lavori di Alan Turing, universalmente riconosciuto come il padre di tale scienza.

Il compito che si assume l'informatica è quello di modellizzare e risolvere problemi, dove la prima abilità è importante almeno quanto la seconda. Per raggiungere questi obiettivi essa usa strumenti chiamati algoritmi, attraverso i quali fornisce una descrizione finita e non ambigua di una sequenza di operazioni che consentono ad un certo agente di risolvere un determinato problema.

Nei programmi di ricerca riguardanti l'IA sono da sempre presenti due filoni d'indagine principali: l'IA guidata dai modelli (IA simbolica) e IA guidata dai dati (IA sub-simbolica).

L'IA simbolica fa largo uso della logica matematica per la rappresentazione della conoscenza in generale e del ragionamento automatico, sia esso di tipo deduttivo che abduttivo, in particolare. Tale programma viene usato in ambito teoretico mediante la dimostrazione automatica dei teoremi (*theorem proving*) e la verifica automatica delle loro dimostrazioni (*proof checking*).

Le applicazioni dell'IA simbolica possono essere così riassunte:

– la *pianificazione automatica*. È questa un'operazione dove una volta assunti i dati della situazione iniziale, un obiettivo da raggiungere e un insieme di vincoli, si richiede alla macchina di costruire un *piano d'azione* (se esiste) che consenta di raggiungere l'obiettivo a partire dalla situazione iniziale e rispettando i vincoli dati. Si vuole produrre una macchina che sia capace di eseguire questo piano d'azione, di monitorarne l'esecuzione e in caso si rilevino situazioni impreviste o indesiderate riformulare un nuovo piano. Normalmente la si usa per l'automazione dei processi di produzione, ma anche per le funzioni che svolgono i diversi satelliti artificiali o sonde spaziali.

– la *robotica collaborativa*. La robotica contemporanea si sta sviluppando nella direzione della robotica collaborativa dove si punta alla collaborazione tra umani e robot per portare a termine un dato obiettivo in ambito industriale o domiciliare. In questo senso la robotica collaborativa combina le potenzialità e le capacità delle persone con quelle dei robot per portare a termine, in maniera semi-automatica, dei compiti che risultano impossibili da realizzare mediante una operazione totalmente automatica. I robot collaborativi (cobot) sono dei robot tipicamente industriali di nuova generazione progettati per lavorare assieme all'essere umano, fianco a fianco e in sicurezza, senza la necessità di barriere o gabbie protettive che separino gli uni dagli altri, ma vi sono applicazioni anche nell'ambito dell'assistenza domiciliare e della domotica.

– il *Business Process Management* (BPM). Sono strumenti informatici che nella gestione dei processi aziendali permettono di migliorare molto la prestazione del

sistema. Si mette a disposizione delle aziende una molteplicità di strumenti per la gestione dei loro processi. Questi sono strumenti capaci di esplicare varie funzioni: a) consentono di scoprire, rappresentare ed analizzare i compiti sottostanti ai processi aziendali (*process mining*); b) consentono di visualizzare i processi aziendali (*diagrams*); c) strumenti per l'automazione dei processi o della gestione dei processi aziendali (*workflow engine*); d) infine strumenti per la simulazione e il testing di processi aziendali.

Più brevemente l'IA sub-simbolica lavora per raggiungere l'obiettivo di imitare, o meglio, mimare il comportamento del cervello umano e la sua complessa rete di neuroni interconnessi. I primi modelli di reti neurali risalgono già agli anni '40 del secolo scorso con gli studi primordiali di neuroscienze computazionali mediante la modellizzazione dei neuroni da parte del neurofisiologo Warren S. McCulloch e del matematico Walter Pitts[7].

I due filoni non sono tra loro separati nel senso che i problemi dati possono essere risolti con uno o l'altro filone. Il caso classico è l'elaborazione del linguaggio naturale dove i primi lavori usarono l'approccio simbolico dove la linguistica computazionale ebbe un grande sviluppo e produsse degli ottimi lavori ma non sempre conclusivi. Successivamente si è seguito l'approccio sub-simbolico fornendo al sistema grandi collezioni di testi scritti e orali prodotti in contesti di comunicazione reale (ad esempio testi della rete) e utilizzando i prodotti per l'apprendimento automatico (*machine learning*) i cui risultati hanno espresso le varie Chat-GPT.

I due programmi di ricerca *IA simbolica* e *IA sub-simbolica* non sono tra loro separati, pur avendo origini diverse sembrano destinati ad interagire tra di loro.

Da sempre abbiamo cercato di far fare alle macchine delle cose che non eravamo in grado di fare, o che non volevamo fare; ciò non ci ha mai spaventato. Oggi la novità è che i sistemi di IA sono in grado di sostituire l'essere umano e, spesso, di migliorarne le prestazioni in compiti ritenuti da sempre di sua competenza esclusiva. Compiti che hanno a che fare con il ragionamento e la memoria.

Autonomia delle macchine, problemi di affidabilità e rischi

Una questione che non possiamo tralasciare è quella relativa all'autonomia delle macchie. Questo è un argomento fondamentale in quanto l'essere umano vorrebbe essere sempre in grado di potersi inserire nel ciclo di comportamento di una macchina e intervenire dicendo «Si» o «No» all'avanzamento del programma. Nei sistemi robotici, e più in generale nei sistemi di IA, non sempre è possibile garan-

7 - W. McCulloch - W. Pitts, *A Logical Calculus of Ideas Immanent in Nervous Activity*, in «Bulletin of Mathematical Biophysics» 5 (1943), pp. 115-133.

tire la condizione che preveda l'interazione col soggetto umano nei processi di decisione (*human in the loop*). Questa impossibilità è dovuta alla velocità operativa della macchina che è molto superiore a quella dell'essere umano nel prendere una decisione.

In questo senso la macchina è autonoma, ma non nel senso che ci sia un'assenza del controllo umano. È bene non confondere l'incapacità umana di prevedere e controllare il comportamento di una macchina con un'attività "creativa" della macchina stessa.

Più una macchina è complessa, e quindi potente, più risulta difficile controllarne totalmente il comportamento. Se vogliamo lasciare la macchina completamente autonoma dobbiamo esser sicuri che anche quei comportamenti che non siamo in grado di poter controllare non siano pericolosi. Il livello di autonomia che siamo disposti ad accettare deve essere commisurato al grado di criticità di certi comportamenti. Si pensi al caso delle auto a guida autonoma, ma anche all'uso dell'IA in ambito militare.

Riguardo il problema dell'affidabilità si pone la questione: «possiamo credere a quello che fa o che dice il mio sistema di IA?». In molti casi, quali le applicazioni finanziarie e in diversi sistemi critici dal punto di vista della sicurezza quali velivoli, centrali nucleari in cui l'intervento deve essere rapido, la condizione *human in the loop* non può essere garantita e quindi soddisfatta. Allora occorre costruire dei sistemi di cui ci si può fidare (*trustworthy AI*). L'IA di cui ci si può fidare è quella macchina che è in grado di spiegare e rendere ragione delle sue conclusioni o previsioni. L'obiettivo del *trustworthy AI*, a oggi, non è purtroppo ancora stato raggiunto. Una possibile strada risolutiva potrebbe essere quella risultante dall'integrazione dei lavori tra l'IA simbolica e l'IA sub-simbolica.

A questo punto non possiamo non interrogarci su quale sarà il carattere qualitativo delle relazioni tra uomo e macchina dopo il pieno avvento dell'innovazione tecnologica dell'IA. Relazioni che potrebbero possedere delle finalità utili ma anche far sorgere manipolazioni del comportamento delle persone.

Già oggi abbiamo in rete decine si siti[8], dove si può programmare la propria ragazza virtuale; un avatar con la quale milioni di utenti intrattengono conversazioni e che a sua volta è capace di relazionarsi con il proprio ideatore a livello quasi umano. Sarebbe un errore ritenere che tali relazioni possano essere considerate dei fenomeni di nicchia quando in questo mercato miliardario le cinque app più popolari sono state scaricate oltre 27 milioni di volte[9]. Questa non è altro che una sin-

8 - Cfr. *Virtual Girlfriend, Ai Girlfriend o iGirl*.

9 - Cfr. il sito *What's the big data*, https://whatsthebigdata.com/ai-girlfriend-statistics/ (accesso: 23 gennaio 2025).

gola quanto allarmante manifestazione di una nuova tipologia di relazione tra uomo e macchina che le persone potranno avere con l'uso di modelli avanzati di IA.

È dunque più che mai lecito chiedersi quale tipo di relazione uomo/macchina è possibile stabilire con modelli di IA, se non altro perché nel mondo vi sono ormai milioni di persone che dicono di sperimentare una relazione empatica con queste macchine, ignorando o omettendo che le risposte ottenute sono frutto di algoritmi. La domanda possiede una ulteriore legittimità poiché questo tipo di relazione diventerà sempre più comune in futuro, con modalità che certamente renderanno sempre più difficile capire se il proprio interlocutore, la persona con la quale si sta intrattenendo una relazione, sia realmente un essere umano, o invece un computer.

Nel suo relazionarsi con le IA, il soggetto umano potrebbe mettere in atto dei comportamenti psicologici particolari che ne corroborano il narcisismo, o che potrebbero generare esperienze solipsistiche, isolandolo in un mondo virtuale nel quale si rispecchia e si riconosce completamente, riducendo progressivamente i legami "reali" con i suoi simili. Nel già noto fenomeno dei «reclusi nella rete» (*hikikomori*) assistiamo al fatto che le relazioni virtuali hanno sostituito progressivamente quelle reali. Se in questo caso le relazioni – seppur virtuali – sono con altri esseri umani, nulla esclude che nel prossimo futuro le relazioni degli *hikikomori* divengano anche con macchine a IA.

Le relazioni con macchine a IA, sotto certi aspetti, possono essere più soddisfacenti rispetto a quelle che si possono avere con i propri interlocutori umani, poiché tali macchine possono empatizzare con il proprio interlocutore, in quanto lo capiscono meglio di altri, ne sanno anticipare i bisogni e ne rafforzano le proprie convinzioni. Uno studio preventivo su come muterà il rapporto tra persona e macchina, grazie all'uso dell'IA, si rivela quindi essenziale sia per comprendere la società di domani sia per sciogliere i molti interrogativi che in noi sorgeranno.

Conclusioni

Ciò che può rendere problematica una innovazione tecnologica non sono solo quei rischi che essa intrinsecamente veicola, ma gioca un ruolo importante anche il contesto culturale e sociale in cui tale innovazione appare e si afferma. Contesto che può essere preparato ad assorbire l'impatto e poi essere capace di una buona ricezione dell'innovazione, oppure ad amplificare quei caratteri negativi che essa ha già in gestazione.

La ricezione sociale delle innovazioni tecnologiche mostra un carattere ambivalente in quanto vi sfociano componenti di natura diversa come quella psicologica e quella pragmatica. Il carattere pratico-pragmatico determina quella serie di questioni che riguardano lo scopo (non certamente disinteressato) del produrre l'innovazione, la varietà di mezzi necessari alla sua creazione (materiali, *know how*) e il

contesto (non sempre ben conosciuto) con le relative conseguenze sociali. Questa tematica schiude tutta una gamma di concetti e giudizi di valore che riguardano la sfera etica. Sfera già in serie difficoltà ai giorni nostri, ma che entra ancor più in difficoltà quando tenta di esprimere i propri giudizi sulle innovazioni tecnologiche proposte da gruppi economici e industriali, dai complessi militari-industriale e dai sistemi politici, tutti soggetti che nel perseguire i propri obiettivi mostrano quotidianamente una certa impermeabilità ai valori etici.

Come ogni innovazione tecnologica anche l'IA potrà aiutare la persona umana in vari aspetti del suo vivere personale e sociale, ma inevitabilmente procurerà anche quei problemi già intrinseci all'innovazione e che verranno amplificati dal contesto culturale segnato dall'era del dominio della tecnica dove non sarà facile reperire quella saggezza e responsabilità necessarie ad un autentico sviluppo dell'umano. Il timore che, tale innovazione tecnologica procuri un ulteriore sviluppo della razionalità funzionale della tecnica, a scapito della razionalità speculativa umana, con il conseguente sfaldamento della capacità delle persone di affrontare questioni sul senso dell'esistenza e sui fini che caratterizzano le nostre azioni, è più che fondato. Nel momento in cui dovremmo promuovere una riflessione etica e una maggiore equità sociale per la rinascita dell'umano ci troveremmo ad affrontare non una qualunque innovazione tecnologica o un gigantesco balzo verso una quarta rivoluzione[10] industriale ma un vero e proprio cambiamento d'epoca senza una saggezza adeguata. Questo perché l'IA segna una nuova e significativa fase nel rapporto dell'umanità con la tecnologia dove i confini tra l'umano e la macchina, tra il naturale e l'artificiale, tra mondo fisico e mondo virtuale diventeranno sempre più opachi e confusi.

Risulterà allora ancora più difficile recuperare quella vita interiore, capace di produrre in noi quella saggezza che sappia unire l'uso cooperativo della razionalità speculativa e della razionalità funzionale, che possa esercitare in modo integrativo le nostre capacità contemplative con quelle tecniche.

In tale contesto – nelle società industriali – occorre mettere a fuoco il problema della ricezione sociale delle innovazioni prodotte dalla tecnica e contemporaneamente una linea di pensiero a sostegno di una strategia eticamente responsabile per le sue applicazioni.

C'è da chiedersi, vista la globalità delle popolazioni che verranno investite dalle innovazioni tecnologie riguardanti l'informatica, se non sia il caso di dare voce alla comunità internazionale che a partire dalla presa di coscienza di responsabilità degli stati sovrani, produca un documento sotto forma di trattato internazionale

10 - L. Floridi, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo,* Raffaello Cortina, Milano 2017.

che possa essere – almeno in alcuni aspetti – giuridicamente vincolante in vista di una appropriata regolamentazione etica delle applicazioni dell'IA nei moltissimi campi della vita sociale[11].

Le attuali organizzazioni internazionali potrebbero coinvolgere il settore privato dell'alta tecnologia nello svolgere un ruolo decisivo per raggiungere accordi multilaterali e avviare iniziative individuali o di gruppi di stati capaci di incoraggiare e sostenere migliori pratiche riguardanti l'IA e stimolare nuovi approcci creativi al problema.

SERGIO RONDINARA
Già Professore ordinario di Epistemologia presso l'Istituto Universitario Sophia (FI)
sergio.rondinara@sophiauniversity.org

11 - L'*AI Act*, il regolamento europeo sull'IA varato dall'Unione Europea sembra un primo piccolo passo che va in questa direzione.

Il ruolo della transdisciplinarità nell'era dell'intelligenza artificiale (generativa)

di Marco Sanchioni

The essay addresses the challenges posed by artificial intelligence to the production of knowledge, highlighting the limits of machines in understanding and generating meaning. Starting from these considerations, transdisciplinarity is proposed as an epistemological and formative paradigm capable of preserving the specificity of human intelligence. Finally, the decisive role of universities in promoting generative knowledge, capable of orienting and transforming, is shown.

Keywords: Transdisciplinarity, Artificial Intelligence, Knowledge, Epistemology, Philosophy of Science

1. Introduzione

L'intelligenza artificiale sta trasformando radicalmente il nostro modo di immaginare il futuro, mettendo in discussione il ruolo stesso dell'essere umano anche in molti ambiti della conoscenza. Se un tempo attribuivamo certe capacità esclusivamente all'intelligenza umana – la scrittura, la musica, la diagnosi medica, la progettazione teorica – oggi vediamo macchine che non solo replicano queste competenze, ma in alcuni casi le eseguono meglio di noi. Basti pensare alla medicina, dove sistemi di intelligenza artificiale come *DeepMind's AlphaFold* hanno rivoluzionato la biologia computazionale, prevedendo la struttura delle proteine con una precisione senza precedenti. Oppure alla musica e alla scrittura: strumenti come *OpenAI's Jukebox* possono generare composizioni musicali indistinguibili da quelle di un essere umano, e modelli come *ChatGPT* producono testi coerenti, creativi e talvolta sorprendenti nella loro capacità di imitare lo stile e il ragionamento umani. Anche nel mondo dell'arte, le IA generative come *DALL·E* e *MidJourney* creano immagini innovative che sfidano i confini della produzione artistica tradizionale. Cose che per secoli abbiamo considerato tipicamente umane sembrano non esserlo più.

L'IA ci supera in efficienza, rapidità e precisione, dimostrando una capacità di elaborazione che, per alcuni compiti, va oltre ciò che un singolo individuo potrebbe mai realizzare. Di fronte a questa rivoluzione, sorge spontanea una domanda essenziale: quale sarà il ruolo dell'essere umano nel futuro? Se le macchine sono sempre più competenti nelle attività che un tempo costituivano il nucleo delle professioni intellettuali e creative, dove si colloca il contributo insostituibile dell'essere umano? Se l'IA è in grado di scrivere saggi, proporre nuove teorie scientifiche e persino comporre sinfonie, cosa rimane esclusivamente nostro?

La domanda è indubbiamente complessa e non intendiamo fornire una risposta esaustiva in questo saggio. Tuttavia, vogliamo proporre una chiave di lettura che si concentri in particolare sul problema della conoscenza. Basarab Nicolescu introduce quello che potremmo definire un cambio di paradigma, individuando, in maniera particolare, una crisi epistemologica legata alla frammentazione del sapere e alla crescente influenza della tecnoscienza. Egli scrive:

> Tenendo presente che l'attuale proliferazione delle discipline accademiche e non accademiche conduce ad una crescita esponenziale del sapere, cosa che rende impossibile lo sguardo globale dell'essere umano; […] tenendo presente che la vita è pesantemente minacciata dalla scienza tecnica trionfante, che non obbedisce che alla temibile logica della efficienza per l'efficienza; […] tutto ciò

considerato, i partecipanti al Primo Congresso Mondiale di Transdisciplinarità (Convento di Arrábida, Portogallo, 2-7 novembre 1994) adottano la presente Carta […][1].

Questo passaggio costituisce il preambolo del Manifesto dell'allora nascente transdisciplinarità, scritto in un'epoca in cui le conseguenze della specializzazione accademica e dell'accelerazione tecnologica erano già evidenti, ma in cui l'intelligenza artificiale non aveva ancora raggiunto il livello attuale.

L'idea di transdisciplinarità, però, non nasce con Nicolescu. Il termine appare per la prima volta nel 1970, nel convegno internazionale *Interdisciplinarity – Teaching and Research Problems in Universities*, organizzato dall'OCSE in collaborazione con il Ministero dell'Istruzione francese e l'Università di Nizza. In quel contesto, Jean Piaget descrive la transdisciplinarità come un livello superiore all'interdisciplinarità, caratterizzato dall'assenza di confini stabili tra le discipline[2]. Negli stessi anni, Erich Jantsch e André Lichnerowicz propongono visioni alternative: Jantsch interpreta la transdisciplinarità come una forma di coordinamento tra discipline e innovazione nel sistema educativo[3], mentre Lichnerowicz la concepisce come una generalizzazione matematica dell'attività scientifica[4]. Queste prospettive, pur contribuendo al dibattito, rimangono legate a un quadro disciplinare chiuso. Sarà solo con Nicolescu, negli anni '90, che la transdisciplinarità assumerà la sua forma attuale, distinguendosi come un nuovo paradigma epistemologico. Come vedremo, essa non si limita alla collaborazione tra discipline, ma integra *livelli di realtà* diversi. Non si esaurisce in un metodo, ma è una forma di conoscenza che coinvolge il soggetto. Non si basa sulla logica classica, ma sul principio del *terzo incluso*, che permette di superare le contraddizioni senza ridurle a un semplice compromesso.

Il preambolo del manifesto sulla transdisciplinarità di Nicolescu risulta oggi più attuale che mai. L'iper-specializzazione del sapere e l'autonomia crescente della tecnoscienza pongono nuove sfide: che tipo di conoscenza è possibile in un mondo in cui le macchine sembrano poter superare l'essere umano nelle sue stesse capa-

1 - B. Nicolescu - E. Morin - L. de Freitas, *La Carta della Transdisciplinarità*, Convento di Arrábida, 6 novembre 1994, trad. it. a cura dell'I.P.E. – Istituto per ricerche ed attività educative, Napoli, pubblicata in https://disf.org/carta-transdisciplinarieta (consultato il 23 maggio 2025). Il testo è tratto dalla versione originale francese inclusa in B. Nicolescu, *La transdisciplinarité. Manifeste*, Éditions du Rocher, Monaco 1996, p. 13.

2 - J. Piaget, *L'épistémologie des relations interdisciplinaires*, in L. Apostel et al., *Interdisciplinarity: Problems of Teaching and Research in Universities*, OCSE, Paris 1972, p. 144.

3 - E. Jantsch, *Inter- and Transdisciplinary University: A Systems Approach to Education and Innovation*, in «Higher Education», 1 (1972), pp. 7-37.

4 - A. Lichnerowicz, *Mathématique et transdisciplinarité*, in «ECON. SOC.», 6/8 (1972), pp. 1497-1509.

cità cognitive? E, più radicalmente, esiste una forma di sapere – non intesa come semplice insieme di conoscenze, ma come modalità epistemica[5] – che l'IA non potrà mai raggiungere? È proprio a partire da questa riflessione che proponiamo di considerare la transdisciplinarità come una delle prospettive più significative per comprendere il ruolo dell'essere umano nell'epoca delle IA generative. Non solo come metodo, ma come *spazio epistemico* in cui il sapere si fa autenticamente generativo, capace, come vedremo in questo saggio, di integrare livelli di realtà diversi, trasformare il soggetto e generare orizzonti di senso.

Un aspetto ulteriore che intendiamo approfondire riguarda il ruolo cruciale delle università in questo nuovo scenario. Nel corso degli ultimi decenni, molte istituzioni accademiche hanno adottato un modello sempre più orientato alla specializzazione, spinte da diverse dinamiche: da un lato, la crescente pressione esercitata dalle logiche di mercato, che tendono a misurare il valore della formazione in termini di impiegabilità e produttività immediata; dall'altro, il legittimo desiderio di offrire un servizio concreto alla società, formando professionisti altamente qualificati in grado di inserirsi rapidamente nel mondo del lavoro. Di conseguenza, l'università ha progressivamente trasformato i propri programmi, frammentando i percorsi formativi in corsi sempre più settoriali e altamente tecnici, focalizzati sull'acquisizione di competenze specifiche piuttosto che su una visione integrata del sapere. Questo approccio ha portato a una sorta di parcellizzazione del sapere accademico, in cui la profondità della conoscenza è stata privilegiata rispetto alla capacità di costruire connessioni tra discipline diverse. Ma di fronte alle trasformazioni radicali introdotte dall'intelligenza artificiale, è ancora sostenibile questo modello? Se fino ad oggi l'obiettivo principale delle università è stato quello di fornire strumenti per l'inserimento lavorativo, ha ancora senso puntare esclusivamente sulla specializzazione, quando le IA stanno dimostrando di poter eccellere proprio in molti di questi compiti specialistici? Forse è arrivato il momento di ripensare il ruolo dell'università, non necessariamente abbandonando la formazione specialistica, ma complementandola con un modello più ampio, capace di affrontare le sfide della nuova rivoluzione tecnologica.

Se la transdisciplinarità, come suggerisce Nicolescu, rappresenta uno spazio in cui la conoscenza diventa autenticamente generativa – cioè capace di trasformare il soggetto e costruire significati nuovi oltre la semplice elaborazione di dati – allora è proprio in questo spazio che le università potrebbero trovare la loro missione per il futuro. Se accetteranno questa sfida, potranno diventare luoghi *generativi*,

5 - Con "modalità epistemica" si intende una forma di accesso alla conoscenza legata non solo ai contenuti acquisiti, ma al modo in cui essi vengono costruiti, vissuti e trasformati all'interno di un'esperienza conoscitiva situata e soggettivamente implicata.

ambienti in cui la formazione non si limiti alla trasmissione di competenze tecniche, ma diventi una vera e propria officina della conoscenza, capace di integrare saperi diversi, stimolare il pensiero critico e offrire un orizzonte di senso in un mondo sempre più dominato dall'efficienza computazionale. In questo senso, l'università non dovrebbe essere soltanto un'istituzione che risponde alle esigenze del mercato, ma un laboratorio per il futuro dell'umanità, in cui il sapere viene continuamente ricreato e rielaborato in modo dinamico e transdisciplinare.

L'articolo è strutturato come segue. Inizieremo introducendo gli elementi essenziali della transdisciplinarità secondo Basarab Nicolescu (sezione 2), soffermandoci su concetti chiave come il *terzo incluso*, l'integrazione di livelli di realtà e il *sapere generativo*. Successivamente (sezione 3), forniremo una panoramica sul funzionamento delle intelligenze artificiali generative, evidenziando come abbiano già superato l'essere umano in molte competenze ultraspecialistiche, concentrandoci principalmente su un loro particolare utilizzo nella ricerca scientifica. A questo punto (sezione 4), ci chiederemo se le IA attuali siano capaci di transdisciplinarità nel senso di Nicolescu, valutandone i possibili limiti. Infine (sezione 5), commenteremo il ruolo delle università in questa transizione, chiedendoci se il modello iper-specialistico sia ancora sufficiente o se debba essere integrato con un approccio transdisciplinare, rendendo l'università un *luogo generativo* di conoscenza per il futuro.

2. Transdisciplinarità e Sapere Generativo

Prima di addentrarci nell'argomento, è importante chiarire a quale concezione di transdisciplinarità ci riferiamo in questo saggio. Il termine, infatti, è stato interpretato in modi diversi, dando origine a due approcci distinti. Da un lato, la visione proposta da Basarab Nicolescu, che ha introdotto una prospettiva radicalmente nuova sulla conoscenza, includendo aspetti epistemologici, etici e metafisici. Dall'altro, l'approccio della cosiddetta "scuola di Zurigo"[6], che ha sviluppato una concezione più pragmatica della transdisciplinarità, concentrata sulla progettazione di soluzioni applicabili a problemi concreti della società contemporanea. La distinzione tra questi due modelli è netta[7]. La transdisciplinarità di Nicolescu si concentra sulla relazione tra diversi livelli di realtà, sul coinvolgimento attivo del soggetto nella costruzione della conoscenza e su una visione unitaria del sapere. La scuola di Zurigo, invece, assume un carattere prevalentemente operativo e meto-

6 - J. Thompson Klein, *Prospects for transdisciplinarity*, in «Futures» 36 (4/2004), pp. 515-526.

7 - S.L.T. McGregor, *The Nicolescuian and Zurich Approaches to Transdisciplinarity*, in «Integral Leadership Review» 15 (3/2015): http://integralleadershipreview.com/13135-616-the-nicolescuian-and-zurich-approaches-to-transdisciplinarity/

dologico, privilegiando l'interfaccia tra scienza, società e tecnologia. Lo stesso Nicolescu ha riconosciuto questa differenza, definendo il proprio approccio come transdisciplinarità teorica, in contrapposizione alla transdisciplinarità fenomenologica come quella elaborata da Michael Gibbons e Helga Nowotny[8], la quale si concentra sulla costruzione di modelli che collegano princìpi teorici con dati sperimentali già osservati, al fine di prevedere ulteriori risultati[9]. Nel contesto di questo saggio, non ci occuperemo della scuola di Zurigo. Il motivo è semplice: la transdisciplinarità metodologica, intesa come strumento di integrazione di conoscenze per risolvere problemi concreti, risulta essere, probabilmente, perfettamente realizzabile da un'intelligenza artificiale. Infatti, se la transdisciplinarità si riducesse a un metodo per combinare dati e discipline diverse in modo efficace per risolvere alcuni problemi concreti, nulla impedirebbe alle IA di praticarla. Ciò che ci interessa, invece, è l'altra forma di transdisciplinarità, quella proposta da Nicolescu, che implica un coinvolgimento profondo del soggetto conoscente e una relazione tra livelli di realtà che va oltre la semplice elaborazione di informazioni. Il nostro obiettivo è mostrare perché questa concezione della transdisciplinarità non sia facilmente accessibile all'intelligenza artificiale e resti un'attività in cui l'essere umano gioca un ruolo centrale.

Un altro chiarimento necessario riguarda il modo in cui ci riferiremo al pensiero di Nicolescu. Non intendiamo offrire una rassegna esaustiva della sua opera, né analizzare ogni aspetto della sua concezione della transdisciplinarità. Il nostro obiettivo è piuttosto evidenziare alcune caratteristiche chiave di questo approccio al sapere, quelle che risultano più rilevanti per le riflessioni che svilupperemo nei capitoli successivi. In particolare, ci concentreremo sugli elementi che permettono di distinguere la transdisciplinarità nicolesciana da altre forme di integrazione dei saperi e che saranno centrali nel nostro confronto con le capacità dell'intelligenza artificiale.

---

8 - M. Gibbons et al., *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, SAGE Publications, London 1994; M. Gibbons - H. Nowotny, *The Potential of Transdisciplinarity*, in J. Thompson Klein et al. (eds.), *Transdisciplinarity: Joint Problem Solving among Science, Technology, and Society*, Springer, Basel-Boston-Berlin-Blrkhauser 2001, pp. 67-80.

9 - In altre parole, pur riconoscendo il valore dell'approccio zurighese, che ha il merito di tradurre la transdisciplinarità in pratiche operative di grande utilità per affrontare problemi complessi, in questo saggio non lo prendiamo in esame in modo approfondito. Il nostro intento, infatti, è esplorare la transdisciplinarità come forma di sapere generativo, e dunque ci concentriamo sulla proposta teorica di Nicolescu, che pone al centro la trasformazione del soggetto, l'interazione con livelli differenti di realtà e la dimensione simbolica ed epistemologica del conoscere. L'approccio zurighese, invece, si sviluppa principalmente come metodologia integrativa tra saperi, priva di una riflessione esplicita sullo statuto ontologico del sapere e sulla sua dimensione trasformativa.

2.1 L'assioma Ontologico

Uno degli elementi centrali della concezione della transdisciplinarità di Nicolescu è ciò che egli chiama assioma ontologico:

> Esistono, nella Natura e nella società, così come nella nostra conoscenza della Natura e della società, diversi livelli di Realtà dell'Oggetto e, corrispondentemente, diversi livelli di Realtà del Soggetto[10].

Questa distinzione tra livelli di realtà dell'oggetto e del soggetto è fondamentale per comprendere la transdisciplinarità come un sapere che non si limita a connettere discipline diverse, ma che riconosce la coesistenza di molteplici livelli del reale, ciascuno caratterizzato dalle proprie leggi e dai propri limiti epistemologici.

Per chiarire questa nozione, Nicolescu distingue tra *Real* e *Reality*. Con *Real*, egli indica ciò che è, indipendentemente dalla nostra esperienza o dalle nostre categorie cognitive, mentre *Reality* è ciò che oppone resistenza nella nostra esperienza, ossia ciò che possiamo conoscere attraverso la nostra interazione con il mondo. Possiamo comprendere questa distinzione riflettendo sul fatto che, a livello microscopico e macroscopico, è estremamente difficile individuare un sistema fisico chiuso. Nel dominio quantistico, la stessa nozione di individualità è messa in crisi da fenomeni come l'entanglement, che rendono molto difficile attribuire un'identità autonoma a un sistema fisico. Nel dominio macroscopico, invece, i sistemi fisici sono sempre aperti, scambiando continuamente energia e informazione con l'ambiente circostante. Questo ci porta a concludere che, da un punto di vista fisico, dovremmo ammettere l'esistenza di un solo sistema davvero chiuso: l'universo nel suo complesso. Tuttavia, per poter fare fisica e costruire teorie scientifiche, siamo costretti a introdurre approssimazioni che ci permettano di trattare certi sistemi *come se* fossero individuali. Si tratta, in realtà, di quasi-individui, entità approssimativamente reali che emergono dal nostro bisogno di modellizzazione. Seguendo questa linea di pensiero, possiamo dire che ogni volta che introduciamo un sistema fisico "chiuso", stiamo operando una semplificazione rispetto alla totalità del reale. Nel linguaggio di Nicolescu, potremmo dire che queste operazioni modellizzanti si collocano entro un certo livello di realtà, inteso come quel dominio in cui l'esperienza e la formalizzazione trovano coerenza attraverso un insieme specifico di leggi e rappresentazioni. Ogni livello di realtà è caratterizzato dalla propria

---

10 - Traduzione mia da: B. Nicolescu, *Methodology of Transdisciplinarity. Levels of Reality, Logic of the Included Middle and Complexity*, in «Transdisciplinary Journal of Engineering & Science» 1 (1/2010), pp. 19-38, qui 24. Testo originale: «There are, in Nature and society and in our knowledge of Nature and society, different levels of Reality of the Object and, correspondingly, different levels of Reality of the Subject».

incompletezza, nel senso che le leggi che lo governano sono solo una parte del complesso di leggi che regolano la totalità del reale[11].

Nicolescu introduce inoltre un concetto fondamentale per comprendere il legame tra questi livelli: la zona di non-resistenza. Egli scrive:

> La zona tra due livelli diversi, e al di là di tutti i livelli, è una zona di non-resistenza alle nostre esperienze, rappresentazioni, descrizioni, immagini e formulazioni matematiche. Questa zona di non-resistenza corrisponde al sacro, a ciò che non si sottomette ad alcuna razionalizzazione[12].

Nel lessico di Nicolescu, razionalizzazione non indica semplicemente l'uso della ragione, ma un processo specifico attraverso cui l'esperienza e la realtà vengono ridotte a modelli formali, concettuali o matematici[13]. È l'operazione tipica della scienza moderna, che tende a eliminare ambiguità e contraddizioni per costruire descrizioni coerenti e prevedibili del mondo. Tra un livello di realtà e l'altro non esiste una transizione continua e graduale, ma una discontinuità ontologica. Nell'approccio di Nicolescu, ogni livello è separato dagli altri da una zona in cui il sapere razionale non può penetrare, una soglia in cui le nostre categorie di descrizione perdono la loro efficacia. Questa zona di non-resistenza non è un vuoto epistemico, ma una dimensione che consente il passaggio da un livello all'altro, permettendo la coesistenza di prospettive apparentemente inconciliabili. Nicolescu identifica questa zona con il sacro, non nel senso religioso del termine, ma come ciò che sfugge alla razionalizzazione completa, ciò che non può essere ridotto a una descrizione formale. In altre parole, la zona di non-resistenza, che egli associa al sacro, resiste a questa riduzione scientifica: non può essere completamente rappresenta-

---

11 - Cfr. V. Fano - M. Sanchioni, *Relational Properties and Relational Quantum Mechanics*, forthcoming, 2025. Si noti che, tuttavia, un livello di realtà, secondo Nicolescu, non coincide con un sistema fisico modellizzato, né con un dominio tecnico di studio. Si tratta piuttosto di un regime ontologico-epistemico, definito da un insieme di leggi irriducibili ad altri livelli e da una specifica relazione tra soggetto e oggetto. I sistemi chiusi o quasi-individui che usiamo nella modellizzazione scientifica operano entro un livello di realtà, ma non lo esauriscono: ne sono, al più, strumenti approssimativi.

12 - B. Nicolescu, *Methodology of Transdisciplinarity. Levels of Reality…*, cit., p. 26, traduzione mia. Testo originale: «The zone between two different levels and beyond all levels is a zone of non-resistance to our experiences, representations, descriptions, images, and mathematical formulations. The zone of non-resistance corresponds to the sacred–to that which does not submit to any rationalization».

13 - Questo uso si situa nella critica alla razionalità riduzionista della modernità, già presente in Werner Heisenberg, che distingue tra l'immagine della natura e la natura stessa (*Physics and Philosophy*, 1958); in Kurt Gödel, che mostra i limiti della formalizzazione logica nei sistemi deduttivi (*On formally undecidable propositions of principia mathematica and related systems*, 1931); e in Edmund Husserl, che denuncia la crisi delle scienze europee per la loro incapacità di fondarsi sull'esperienza vissuta (*Die Krisis der europäischen Wissenschaften*, 1936).

ta, formalizzata o spiegata, e proprio per questo rende possibile il passaggio tra diversi livelli di realtà.

Un altro aspetto cruciale della transdisciplinarità di Nicolescu è il ruolo del soggetto nella conoscenza. Egli afferma:

> La totalità delle leggi non esaurisce l'intera Realtà; dobbiamo anche considerare il Soggetto e la sua interazione con l'Oggetto[14].

La conoscenza, nella prospettiva transdisciplinare di Nicolescu, non è un processo meramente oggettivo, né può essere pensata come un atto meccanico di acquisizione di informazioni. Essa implica sempre una relazione viva e dinamica tra soggetto e oggetto. In questa relazione gioca un ruolo decisivo la zona di non-resistenza, ovvero quello spazio che sfugge a ogni riduzione concettuale, che non può essere formalizzato né razionalizzato, e che rende possibile l'emergere della conoscenza. Proprio qui entra in gioco l'*Hidden Third* (Terzo Nascosto), concetto che Nicolescu introduce per superare la dicotomia classica soggetto/oggetto. Il Terzo Nascosto non è un'entità "tra" soggetto e oggetto, come se fosse un ponte neutro o una mediazione esterna, ma è l'elemento relazionale che li mette in comunicazione preservandone la distinzione. È, potremmo dire, la condizione trascendentale del conoscere, ciò che permette l'accesso ai diversi livelli di realtà, e che dà coerenza al processo conoscitivo mantenendone l'apertura. Questa struttura ternaria – soggetto, oggetto, Terzo Nascosto – non è una metafora, ma una proposta epistemologica concreta, che si oppone alla logica bivalente e al realismo ingenuo della scienza classica. Il Terzo Nascosto non è un accessorio, ma ciò che permette l'insorgere stesso del sapere come esperienza trasformativa: una conoscenza che non lascia il soggetto invariato, ma lo trasforma nel processo stesso del conoscere.

Se assumiamo questa prospettiva, non basta forse dire che l'IA non ha soggettività: dobbiamo riconoscere che, senza la presenza del Terzo Nascosto, non esiste neppure il processo conoscitivo in senso transdisciplinare. L'IA, per quanto possa elaborare dati, riconoscere pattern e simulare linguaggio, non partecipa al campo relazionale che genera la conoscenza, perché non possiede livelli di realtà propri, non ha intenzionalità, non ha esperienza. In questo senso, il Terzo Nascosto è strettamente legato al tema dell'intenzionalità: è ciò che rende possibile una coscienza orientata al mondo, capace non solo di rappresentare, ma di essere trasformata dalla relazione con ciò che conosce. Senza questa dimensione relazionale e trasformativa, la conoscenza si riduce a mera informazione – e in questo senso l'IA può accumulare informazione, ma non accedere alla conoscenza come atto pienamente umano. Il punto non è semplicemente che l'IA non ha un "io": è che senza

14 - B. Nicolescu, *Methodology of Transdisciplinarity. Levels of Reality*…, cit., p. 26.

un soggetto che conosce e si trasforma conoscendo, senza l'elemento relazionale che apre e connette i livelli della realtà, nella prospettiva di Nicolescu non c'è vera conoscenza, ma solo calcolo. Ed è proprio questa distinzione a costituire il cuore del sapere transdisciplinare.

2.2 L'assioma del Terzo Incluso

Il secondo assioma della transdisciplinarità, formulato da Nicolescu, mette in discussione la validità assoluta della logica classica. In generale, la logica classica si fonda su tre assiomi fondamentali, che possiamo riassumere così:

- *Assioma di identità: ogni entità è identica a sé stessa.* $\forall x\,(x = x)$
- *Assioma di non-contraddizione: un'affermazione e la sua negazione non possono essere entrambe vere contemporaneamente.* $\forall x \neg \{ P(x) \land \neg P(x)\}$
- *Assioma del terzo escluso: non esiste un terzo valore di verità tra un'affermazione e la sua negazione.* $\forall x \neg \{ P(x) \lor \neg P(x)\}$

Nicolescu contesta in particolare il principio del terzo escluso, ovvero l'idea che tra A e non-A non possa esistere un terzo termine. Egli propone, invece, l'esistenza di un *terzo incluso* (T), che è contemporaneamente A e non-A, ma situato a un livello di realtà differente. Per rappresentare visivamente la logica del terzo incluso, Nicolescu propone un triangolo in cui i tre vertici corrispondono ai termini A, non-A e T. I vertici A e non-A appartengono allo stesso livello di realtà, dove appaiono distinti e in contraddizione tra loro. In questo contesto, il principio di non-contraddizione resta valido: A non può essere allo stesso tempo non-A. Tuttavia, il terzo vertice, T, si trova su un altro livello di realtà. A questo nuovo livello, la separazione tra A e non-A si dissolve, e ciò che prima appariva contraddittorio si rivela come due aspetti di una realtà più ampia e coerente. Nicolescu esprime questa idea dicendo che il terzo incluso agisce su un altro livello di realtà, in cui «ciò che appare separato è in realtà unito, e ciò che appare contraddittorio è percepito come non-contraddittorio»[15]. In questo senso, T non è semplicemente un compromesso tra A e non-A, ma un principio che trascende il loro conflitto, rivelando un nuovo ordine in cui entrambi possono coesistere senza escludersi a vicenda.

Nicolescu sottolinea che questa conciliazione non è definitiva, ma è destinata a essere continuamente superata. Ogni volta che un nuovo livello di realtà viene inte-

15 - Ibid., p. 30, traduzione mia.

grato nella nostra conoscenza, emergono nuove contraddizioni e nuovi paradossi, che a loro volta richiedono un ulteriore livello di integrazione. Questo implica che la conoscenza non potrà mai essere chiusa o completa, ma sarà sempre aperta a nuovi sviluppi. In altre parole, la logica del terzo incluso non abolisce la logica classica, ma ne limita il campo di applicazione, riconoscendo che essa rimane valida all'interno di un singolo livello di realtà, ma diventa inadeguata quando si considerano fenomeni che si estendono su più livelli. Più che un semplice strumento formale, il terzo incluso è un principio di trasformazione, un passaggio tra livelli di realtà che non coinvolge solo il pensiero, ma tutte le dimensioni dell'essere umano.

2.3 L'assioma della complessità e il sapere generativo

La logica del terzo incluso, strettamente connessa alla nozione di livelli di realtà, ci permette di introdurre un concetto di *complessità* che non è semplicemente tecnica, ma strutturale e generativa. Nicolescu distingue, infatti, tra complessità orizzontale e complessità verticale. La prima si riferisce a un'unica dimensione del reale ed è il tipo di complessità che domina il sapere tecnico-scientifico: l'insieme di interazioni, variabili e fattori che rendono un sistema difficile da modellizzare ma sempre interno a un unico livello di realtà. La complessità verticale, invece, è una forma di complessità che attraversa i livelli di realtà e permette di cogliere connessioni che non sarebbero visibili restando confinati in una prospettiva unidimensionale[16].

Edgar Morin, nel suo lavoro sulla complessità, ha sottolineato che il pensiero complesso non è soltanto un ampliamento della conoscenza esistente, ma un cambiamento nel modo di pensare stesso. La complessità, come emerge nella sua riflessione, non si limita ad accumulare informazioni, ma implica un'intelligenza capace di integrare, di vedere l'unità nel molteplice e di riconoscere le relazioni tra fenomeni apparentemente distanti[17]. Questo permette di pensare insieme il tutto e le parti, l'ordine e il disordine, l'uno e il molteplice. Se la complessità è la condizione oggettiva di un mondo fatto di relazioni multiple, interdipendenze, instabilità e livelli di realtà intrecciati, allora il pensiero complesso può essere inteso come la risposta epistemica adeguata a tale condizione. Non si tratta di accumulare dati o ampliare modelli, ma di sviluppare una forma di pensiero capace di riconnettere, attraversare e generare senso. In questo quadro, la complessità diventa terreno fertile per la generatività del sapere: non solo perché impone nuove domande, ma

16 - B. Nicolescu, *Methodology of Transdisciplinarity*, in «World Futures» 70 (3-4/2014), pp. 186-199.

17 - E. Morin - G. Gembillo - A. Anselmo (a cura di), *La sfida della complessità*, Le Lettere, Firenze 2017, pp. 1-79.

perché stimola la costruzione di relazioni tra elementi prima separati o persino contraddittori. È qui che interviene la logica del terzo incluso proposta da Nicolescu: laddove la logica classica vede opposizioni irriducibili, la logica della complessità verticale consente di attraversare i livelli di realtà, producendo configurazioni cognitive nuove e imprevedibili. In altre parole, non è la complessità in sé a generare sapere, ma il tipo di pensiero che sa accoglierla senza ridurla – un pensiero che si lascia trasformare dal reale e che, nel processo stesso del conoscere, trasforma anche il soggetto.

Il sapere generativo è, in questo senso, un sapere che genera significato, che trasforma chi lo esercita e che permette di vedere connessioni inaspettate. È un sapere che non si accontenta di lavorare entro un sistema dato, entro un unico livello di realtà, ma è capace di uscire dalla scatola, rompendo i confini di una logica puramente deterministica. Secondo Morin, il pensiero complesso è quello che affronta l'incertezza senza eliminarla, che abbraccia il paradosso senza ridurlo, che riconosce l'unità nella diversità e la diversità nell'unità[18]. In questo senso, la conoscenza complessa è capace di affrontare l'incertezza, e solo accettando la complessità come parte integrante del reale possiamo arrivare a un sapere che sia veramente creativo e trasformativo. Da questa prospettiva, il sapere generativo non è semplicemente la capacità di elaborare nuove informazioni, ma la capacità di trasformare l'esperienza del sapere stesso. Non è un sapere accumulativo, ma un sapere che si rinnova nella relazione tra soggetto e oggetto, tra livelli di realtà diversi, tra ciò che sembra separato e ciò che invece si scopre connesso.

Interpretando queste riflessioni nella chiave di lettura della transdisciplinarità, potremmo dire che Nicolescu collega questa visione al ruolo del T*erzo Nascosto* (*Hidden Third*), che non è solo un principio logico ma un principio di trasformazione. L'*Hidden Third* agisce come mediatore tra livelli di realtà, che non vanno intesi come livelli disciplinari o semplici registri linguistici, ma come strati ontologicamente distinti del reale, ciascuno caratterizzato da proprie leggi, da una forma specifica di resistenza all'esperienza, e da una particolare relazione soggetto-oggetto[19]. È questo attraversamento dei livelli – non l'accumulazione di informazioni o l'integrazione tra discipline – a rendere possibile la generazione di significati e valori. In questa prospettiva, il sapere generativo non nasce dalla mera articolazione tecnica del sapere, ma dall'incontro tra complessità verticale, esperienza vissuta e trasformazione del soggetto.

---

18 - E. Morin - M. Corbani. *Introduzione al pensiero complesso*, Sperling & Kupfer, Milano 1993.

19 - Per un approfondimento sul ruolo dei livelli di realtà, si veda anche V. Fano - M. Sanchioni, *Relational Properties and Relational Quantum Mechanics*, cit.

> La complessità transdisciplinare non è un sapere chiuso, ma un sapere aperto, che si rinnova nel passaggio tra livelli di realtà e nel riconoscimento della loro interconnessione. Questo sapere non si limita a descrivere il mondo, ma lo trasforma[20].

Se la complessità verticale è il terreno fertile della generatività, allora il sapere generativo è ciò che permette di pensare oltre il dato, oltre il determinato, oltre il calcolabile. La, così potremmo definirla, "vera" creatività – la capacità di vedere possibilità inaspettate – non nasce dall'elaborazione computazionale della complessità orizzontale, ma dall'immersione nel dinamismo della complessità verticale. È forse in questo spazio che si manifesta la qualità irriducibile dell'intelligenza umana: l'IA può certamente analizzare la complessità in termini tecnici, ma la dimensione generativa del sapere – intesa come creazione di nuovi livelli di realtà, di significati, di senso – richiede un soggetto capace di entrare in relazione con il reale e di trasformarsi nel processo conoscitivo stesso. Non si tratta solo di funzioni che l'IA non può eseguire, ma di una modalità dell'essere nel sapere che eccede ogni computazione.

3. Intelligenza Artificiale e il Sapere Scientifico

Ci addentriamo ora nel tema centrale di questo saggio, ovvero lo status dell'intelligenza artificiale rispetto a questi temi. Quando parliamo di IA "intelligente", cosa intendiamo esattamente? Il concetto stesso di intelligenza, quando viene applicato all'intelligenza artificiale, non è univoco né pienamente omologo a quello utilizzato per descrivere l'intelligenza umana. Si tratta piuttosto di un uso funzionale o operativo del termine, riferito alla capacità di risolvere problemi, apprendere da dati e generare output (orizzontalmente) complessi. Con l'evoluzione tecnologica, questa accezione si è ampliata, tanto che alcune prestazioni delle macchine superano oggi quelle umane in compiti specifici. Tuttavia, ciò non implica necessariamente una forma di intelligenza comparabile a quella umana, né una comprensione profonda del contesto in cui questi compiti sono svolti.

In questo scenario, un ambito in cui l'IA sta esercitando un'influenza crescente è quello della ricerca scientifica. Le tecniche di machine learning stanno modificando in profondità il rapporto tra teoria ed esperienza, tra soggetto conoscente e dati osservabili, fino a ridefinire alcuni aspetti fondamentali della costruzione del sapere scientifico. L'impiego dell'IA nella ricerca scientifica è cresciuto esponenzialmente negli ultimi anni, modificando profondamente le modalità di produzione e validazione della conoscenza. Fino a poco tempo fa, attività come la formulazione di ipotesi,

20 - B. Nicolescu, *La transdisciplinarité*, cit.

la progettazione di esperimenti, l'analisi dei dati e la costruzione di modelli teorici erano considerate attività strettamente umane, basate su creatività, esperienza e intuizione scientifica. Con l'avvento del *Machine Learning*, tuttavia, queste funzioni sono state automatizzate e ottimizzate, rendendo il sapere scientifico sempre più ibrido, un intreccio tra cognizione umana e calcolo artificiale. I moderni algoritmi di *machine learning* operano attraverso modelli statistici avanzati, che apprendono da dati storici e sperimentali per individuare pattern, correlazioni e dinamiche sottostanti ai fenomeni osservati. A differenza dei modelli teorici tradizionali, che si basano su leggi dedotte a partire da principi fisici e matematici, il machine learning lavora con un approccio *data-driven*, in cui la conoscenza emerge direttamente dall'analisi dei dati, senza la necessità di un'interpretazione teorica esplicita[21]. Questa differenza è particolarmente evidente nelle tecniche più utilizzate nella ricerca scientifica. Modelli di apprendimento supervisionato (come *XGBoost*, *Random Forests*, *Deep Neural Networks*) vengono addestrati su dataset etichettati per predire risultati e classificare dati sperimentali, migliorando l'affidabilità dell'analisi. Inoltre, l'apprendimento non supervisionato (come *clustering* e *PCA*) consente di scoprire strutture e relazioni latenti nei dati, facilitando la formulazione di nuove ipotesi. Esistono anche tecniche di *reinforcement learning*, sempre più utilizzate in ambito sperimentale, che permettono di ottimizzare iterativamente i parametri di un esperimento, adattandosi dinamicamente ai risultati ottenuti.

L'applicazione di queste tecniche ha avuto un impatto significativo in diversi settori della ricerca. Un caso emblematico è quello esposto da Vera Matarese, che riguarda una ricerca specifica del Brookhaven National Laboratory, dove l'IA è stata impiegata per ottimizzare esperimenti nella fisica degli ioni. Qui, due algoritmi specifici – *GPTune* e *XGBoost* – hanno dimostrato come il *machine learning* possa migliorare radicalmente la precisione e l'efficienza della ricerca sperimentale[22]. Infatti, uno dei problemi più complessi nella ricerca sperimentale è la calibrazione dei parametri: molte esperienze scientifiche dipendono dalla regolazione fine di variabili come temperature, pressioni, intensità di campi magnetici ed elettrostatici. Tradizionalmente, questa ottimizzazione avveniva attraverso prove ed errori, con gli scienziati che regolavano manualmente i parametri per ottenere risultati ottimali. *GPTune*, un algoritmo basato su *Gaussian Process Regression* e *Transfer Learning*, ha permesso di rendere questo processo automatico e adattivo. Il sistema apprende

21 - Cfr. H. Wang et al., *Scientific discovery in the age of artificial intelligence*, in «Nature», vol. 620, n. 7972 (2023), pp. 47-60.

22 - Cfr. V. Matarese, *Virtù intellettuali e intelligenza artificiale: il machine learning può fornire una comprensione scientifica del mondo?*, in «Sophia. Ricerche su i fondamenti e la correlazione dei saperi» XVII (1/2025).

quali combinazioni di parametri portano ai migliori risultati e suggerisce nuove configurazioni sperimentali, adattandosi dinamicamente ai cambiamenti nelle condizioni operative. Questa tecnologia ha permesso di ottimizzare esperimenti che altrimenti avrebbero richiesto settimane o mesi di regolazioni manuali, migliorando l'efficienza del laboratorio e aumentando la precisione dei dati raccolti.

L'adozione del *machine learning* nella ricerca scientifica ha modificato radicalmente il rapporto tra teoria ed esperienza, rendendo il sapere sempre più basato sull'analisi computazionale dei dati. Gli algoritmi di IA non sono più semplici strumenti di supporto, ma elementi determinanti nella gestione della complessità sperimentale e nell'ottimizzazione delle procedure scientifiche. Il fatto che molte attività tradizionalmente affidate all'intelligenza umana – come la selezione di parametri, l'identificazione di correlazioni e l'estrazione di modelli predittivi – possano oggi essere eseguite in modo più efficace da sistemi di apprendimento automatico suggerisce che esiste una forma di sapere in cui l'IA è non solo utile, ma essenziale. Questa dimensione del sapere non si basa sulla costruzione concettuale, ma sulla capacità di processare, organizzare e sfruttare grandi quantità di dati per generare analisi affidabili e migliorare la capacità predittiva della scienza. In questo senso, l'IA è determinante: il suo contributo non è semplicemente quello di accelerare le procedure esistenti, ma di ridefinire il modo in cui l'informazione sperimentale viene trattata e valorizzata. La scienza contemporanea è sempre più dipendente dalla gestione di grandi dataset, e senza il supporto del *machine learning* molte delle scoperte attuali non sarebbero nemmeno concepibili. Forse questo cambiamento non riguarda solo gli strumenti della ricerca, ma anche la figura stessa dello scienziato e il suo ruolo nella pratica sperimentale. Se l'IA riesce a svolgere con maggiore efficienza compiti tradizionalmente affidati agli scienziati – dall'ottimizzazione degli esperimenti all'analisi avanzata dei dati – saranno necessarie competenze diverse per interpretare e guidare il progresso scientifico. Gli scienziati dovranno sempre più interfacciarsi con sistemi intelligenti, comprendere le logiche del *machine learning* e sviluppare nuovi modi per integrare il contributo computazionale nel processo di ricerca.

Possiamo quindi concludere che l'IA ha assunto un ruolo centrale in una forma di sapere tecnico e operativo, in cui la conoscenza non è prodotta attraverso intuizioni concettuali, ma attraverso l'elaborazione efficiente di schemi complessi. Se il sapere scientifico si evolve sempre più nella direzione della modellizzazione e dell'analisi computazionale, allora l'IA non è solo un mezzo per potenziare il lavoro umano, ma una componente fondamentale della costruzione della conoscenza nella sua dimensione tecnica. La questione che rimane aperta è se questo paradigma sia sufficiente per una scienza che non si limiti alla previsione e all'ottimizzazione, ma che continui a interrogarsi sulla natura stessa della comprensione.

4. Intelligenza Artificiale e il Sapere Generativo

A questo punto, possiamo porci una domanda fondamentale: qual è il ruolo dell'essere umano nella produzione di conoscenza nell'era dell'intelligenza artificiale? Il capitolo precedente ha mostrato che l'IA non è più soltanto uno strumento di supporto, ma un attore sempre più determinante nella gestione e nell'organizzazione del sapere tecnico e operativo. In molti ambiti, essa è più efficiente, più rapida, e spesso più precisa dell'intelligenza umana. Ma questo significa che il sapere umano è destinato a diventare superfluo? A prima vista, la risposta potrebbe sembrare affermativa. L'IA non solo eccelle nei compiti computazionali, ma riesce a individuare correlazioni, ottimizzare esperimenti, generare testi e persino simulare strutture linguistiche complesse. Potrebbe perfino apparire in grado di riprodurre forme di sapere che coinvolgono più livelli disciplinari o concettuali[23].

La tesi che vorrei sostenere in questa sezione è che la transdisciplinarità, nel senso proposto da Basarab Nicolescu, costituisce non solo uno "spazio rimasto" per l'intelligenza umana, ma un compito essenziale e irriducibile: è ciò che rende il soggetto umano protagonista attivo nella produzione del sapere. Se la conoscenza scientifica contemporanea tende a frammentarsi in modelli specialistici e data-driven, la transdisciplinarità rappresenta la capacità specificamente umana di attraversare i livelli della realtà, riconoscerne la discontinuità, e ricomporli in una visione complessa e generativa del mondo.

Uno degli argomenti fondamentali per distinguere il sapere umano da quello artificiale riguarda la differenza tra *correlare* e *comprendere*. L'intelligenza artificiale, specialmente nei suoi sviluppi basati sul *Machine Learning*, eccelle nella capacità di individuare pattern e connessioni tra dati. Gli algoritmi più avanzati riescono a riconoscere strutture nei dataset che sarebbero difficilmente accessibili all'analisi umana, permettendo previsioni accurate e ottimizzazioni sofisticate. Tuttavia, il fatto che l'IA trovi correlazioni non significa che comprenda il fenomeno che sta analizzando. È innegabile che essa sia in grado di produrre testi e modelli che rappresentano, anche in modo articolato, un quadro concettuale ampio e strutturato, talvolta evocando un'interazione tra differenti livelli del sapere. Ma questa rappresentazione resta simbolica e derivativa: non nasce da un'esperienza vissuta, né implica un atto intenzionale di comprensione.

Nel senso nicolesciano della transdisciplinarità, comprendere significa attraversare e trasformare i livelli di realtà, generando significato a partire da una parteci-

23 - Si noti che, tuttavia, questa riproduzione avviene, come vedremo in questo capitolo, solo sul piano simbolico: l'IA può generare strutture che rappresentano l'intelligenza, ma non possiede l'esperienza vissuta del sapere, né la capacità di costruire una visione unitaria del mondo che nasca dall'integrazione trasformativa di livelli di realtà diversi.

pazione soggettiva al processo conoscitivo. In questo senso, l'IA può generare output che imitano il risultato simbolico della comprensione, ma non può accedere alla comprensione stessa come atto generativo e trasformativo del soggetto. In altre parole, l'IA, per quanto potente nell'analisi dei dati, non produce nuovi livelli di realtà, ma opera sempre entro le strutture esistenti, elaborando informazioni già disponibili senza introdurre prospettive radicalmente nuove. Questa distinzione è cruciale. Un sistema di intelligenza artificiale può, per esempio, individuare correlazioni tra parametri sperimentali e risultati osservati, ma non genera una teoria che spieghi perché queste correlazioni esistano o quale sia la loro implicazione a livelli più profondi. La scienza umana, invece, non si limita a descrivere le relazioni tra i fenomeni, ma cerca di interpretarle e comprenderle in un quadro più ampio, spesso ridefinendo il concetto stesso di realtà[24].

Prendiamo l'esempio della meccanica quantistica. Gli esperimenti del primo Novecento non solo hanno portato a nuove scoperte, ma hanno ridefinito il concetto stesso di realtà fisica, introducendo nozioni come la sovrapposizione di stati e l'entanglement. Questo processo non è stato semplicemente un'analisi di dati, ma una trasformazione epistemica, un salto concettuale che ha modificato il nostro modo di concepire il mondo fisico. L'intelligenza artificiale, invece, anche nei suoi utilizzi più sofisticati, non introduce simili discontinuità epistemiche. È in grado di generare testi che mettono in relazione domini concettualmente distanti – ad esempio, traducendo il linguaggio della fisica quantistica in metafore per descrivere fenomeni politici o sociali – ma resta confinata a operazioni di ricombinazione simbolica entro strutture cognitive predefinite. Anche le sue scoperte più avanzate, per quanto sorprendenti, non danno luogo a nuovi paradigmi, ma si inseriscono entro modelli costruiti e interpretati dall'essere umano. Se seguiamo Nicolescu e la sua distinzione tra i livelli di realtà, possiamo dire che il sapere umano è caratterizzato dalla capacità di attraversare questi livelli e di crearne di nuovi, mentre l'IA elabora informazioni all'interno di configurazioni date, anche se appartenenti a livelli di realtà diversi. Questo rende il suo contributo essenziale nel dominio tecnico e operativo, dove la funzionalità e l'efficienza guidano la costruzione del sapere.

Tuttavia, l'IA resta inadeguata per una conoscenza che sia realmente generativa, nel senso di capace di produrre trasformazioni epistemiche e nuovi orizzonti di

24 - È importante distinguere tra la capacità dell'IA di generare testi che rappresentano un quadro concettuale ampio – per esempio integrando riferimenti disciplinari diversi, simulando una prospettiva sintetica o articolando argomentazioni complesse – e la capacità di costruire realmente tale quadro attraverso una comprensione vissuta. L'IA, infatti, elabora informazioni simboliche su base statistica e modellistica, producendo testi che imitano la struttura del pensiero umano. Tuttavia, questa produzione non implica un atto intenzionale, né un coinvolgimento epistemico: non vi è coscienza del significato né trasformazione del soggetto, elementi che restano costitutivi del sapere umano.

senso. L'essere umano, quindi, rimane imprescindibile non solo per la sua creatività, ma per la sua capacità di superare la mera funzionalità simbolica, costruendo visioni che trasformano le strutture stesse del pensiero e reinterpretano il reale secondo logiche che l'IA non può né anticipare né fondare autonomamente.

Un secondo motivo per cui l'intelligenza artificiale non può essere agente di transdisciplinarità nel senso di Nicolescu riguarda la sua struttura logica. L'IA, per quanto avanzata, è costruita su sistemi computazionali che operano entro logiche formalmente coerenti, in cui il principio di non contraddizione rimane comunque operante come vincolo strutturale[25]. Questo la rende estremamente potente nell'analisi strutturata dei dati, ma al tempo stesso incapace di navigare livelli di realtà differenti, specialmente quando emergono contraddizioni apparenti tra essi. La transdisciplinarità di Nicolescu si fonda infatti sul superamento della logica classica bivalente. Nel paradigma classico, ogni proposizione è o vera o falsa, e il principio di non contraddizione stabilisce che una proposizione e la sua negazione non possono essere entrambe vere contemporaneamente:

«Per ogni x, non è vero che P(x) e non P(x) siano entrambe vere»:

$\forall x \neg \{ P(x) \land \neg P(x)\}$

Questa struttura logica, come abbiamo visto nella seconda sezione di questo saggio, funziona all'interno di un livello di realtà ben definito. Tuttavia, uno dei punti centrali della transdisciplinarità è che i livelli di realtà non sono isolati, ma interagiscono, e ciò che appare contraddittorio a un certo livello può essere superato in un altro livello. Per Nicolescu, questa dinamica è descritta dal principio del terzo incluso, secondo cui esiste un terzo elemento T che è contemporaneamente A e non-A, ma situato su un altro livello di realtà:

«Esiste un T tale che T = A e T = non-A»: $\exists T (T = A \land T = \neg A)$

Questa struttura non è un paradosso, ma un'apertura verso una logica transdisciplinare, in cui la conoscenza non è semplicemente la coerenza interna di un livello, ma la capacità di attraversare livelli e riconciliare apparenti contraddizioni. Esempi di questa struttura emergono con forza nella fisica quantistica: un elettrone può comportarsi sia come onda che come particella, a seconda del contesto speri-

25 - È vero che alcuni modelli di IA utilizzano logiche non classiche, come la *fuzzy logic*, nelle quali i valori di verità non sono binari ma continui. Tuttavia, tali logiche servono principalmente a gestire incertezze e sfumature nei dati, e non comportano una reale coesistenza di contrari su piani ontologicamente distinti. In particolare, la *fuzzy logic* non introduce un principio epistemologico capace di attraversare livelli di realtà come il *terzo incluso* nicolesciano.

mentale. Questa doppia natura, lungi dall'essere un errore logico, indica che siamo in presenza di una realtà che sfugge alla logica classica, e che può essere colta solo attraverso descrizioni complementari. Tuttavia, nella prospettiva nicolesciana, tale complementarità può essere interpretata come segnale dell'esistenza di livelli di realtà differenti – non semplicemente diversi modi soggettivi di descrivere un fenomeno, ma strati ontologicamente distinti che richiedono modalità conoscitive altrettanto differenziate[26]. L'IA, invece, è costruita su logiche computazionali chiuse e consistenti, in cui le contraddizioni devono essere eliminate o risolte. Anche nei modelli più avanzati, basati su reti neurali profonde, il funzionamento si basa su ottimizzazione e riduzione dell'ambiguità, piuttosto che su una reale capacità di gestire la tensione tra opposti. Le architetture attuali sono progettate per trovare la risposta più probabile, non per accogliere simultaneamente prospettive contraddittorie. Questo significa che, per come è strutturata, l'IA non è adatta a navigare tra livelli di realtà, proprio perché non ha la capacità di tenere dentro le contraddizioni e riconoscerle come parte integrante di un quadro più ampio. La transdisciplinarità richiede un tipo di intelligenza che sia capace di relativizzare il principio di non contraddizione all'interno di un livello di realtà, per aprirsi alla possibilità che ciò che appare contraddittorio possa trovare coerenza a un livello più ampio. Questa operazione non è compatibile con i sistemi computazionali attuali, che restano confinati in una logica di ottimizzazione entro un livello dato. Se la conoscenza transdisciplinare si basa sull'abilità di tenere insieme opposizioni apparenti e di riconfigurare il pensiero alla luce di nuovi livelli di realtà, allora l'IA, per sua costruzione, non può praticarla. Il suo contributo rimane confinato entro schemi logici chiusi e livelli di realtà predefiniti, mentre la vera transdisciplinarità resta un compito specificamente umano, perché implica un attraversamento creativo e non puramente computazionale del sapere.

Un ulteriore elemento che distingue l'intelligenza artificiale dal sapere transdisciplinare è l'assenza di un rapporto soggetto-oggetto nel suo processo di elaborazione delle informazioni. Se consideriamo la conoscenza come un atto in cui un soggetto entra in relazione con un oggetto, costruendo progressivamente significati e trasformando sé stesso attraverso il processo di apprendimento, allora è evidente che l'IA non può realmente conoscere nel senso umano del termine, perché

26 - Alcuni possono interpretare questa dualità onda/particella come una questione epistemologica: a seconda del tipo di esperimento, emergono proprietà diverse, ma non contraddittorie, perché legate a differenti contesti di misura. Nicolescu, invece, propone una lettura ontologica della complessità quantistica: la coesistenza di aspetti incompatibili a un dato livello di descrizione è segno della presenza di livelli di realtà distinti, ciascuno con proprie leggi e modalità di accesso. In questo senso, ciò che nella fisica quantistica appare come complementarità epistemica può essere letto, in chiave transdisciplinare, come discontinuità ontologica. Cfr. B. Nicolescu, *Methodology of Transdisciplinarity*, cit.

non esiste un soggetto che apprende, ma solo un algoritmo che calcola. Nella visione di Basarab Nicolescu, la transdisciplinarità non è solo una metodologia per integrare saperi diversi, ma implica un'interazione profonda tra il soggetto conoscente e i livelli di realtà, mediata dall'*Hidden Third*, quella zona di non-resistenza che connette soggetto e oggetto permettendo una trasformazione reciproca. Per Nicolescu, la conoscenza transdisciplinare è quindi un processo trasformativo, che non riguarda solo l'accumulo di dati, ma la modificazione stessa del soggetto epistemico. Chi conosce, nella prospettiva transdisciplinare, non resta mai identico a sé stesso: il sapere non è un semplice contenuto da immagazzinare, ma un'esperienza che modifica il modo in cui il soggetto percepisce, comprende e si relaziona al reale[27]. L'IA, al contrario, non ha un soggetto conoscente. Non esiste un "io" che si confronta con il sapere, ma solo un sistema computazionale che elabora pattern e produce output sulla base di configurazioni statistiche. Si può certamente dire che l'IA genera rappresentazioni simboliche complesse, e in molti casi produce testi o immagini che simulano i risultati di un processo conoscitivo. Tuttavia, ciò non equivale a un'esperienza del conoscere. Non c'è, nel suo funzionamento, alcuna intenzionalità, alcuna relazione con un mondo vissuto, alcun passaggio dalla latenza del non-sapere alla trasformazione che accompagna l'acquisizione di significato. In questo senso, dire che l'IA "produce conoscenza" è lecito solo in senso operativo: essa genera strutture che possono essere riconosciute come conoscenza da un soggetto esterno, ma non conosce nel senso forte del termine. La conoscenza, nel quadro transdisciplinare qui adottato, è un processo incarnato e situato, che coinvolge il soggetto nella sua interezza: non solo come mente logico-calcolante, ma come coscienza vissuta, attraversata da esperienza, interpretazione e trasformazione. È per questo che, come discusso nelle sezioni precedenti, Nicolescu lega la conoscenza transdisciplinare all'"interazione tra livelli di realtà" e alla mediazione dell'*Hidden Third,* cioè a una dimensione relazionale che coinvolge la totalità dell'essere umano. Questa totalità non è un dato metafisico, ma indica l'impossibilità di ridurre il conoscere a un'unica funzione cognitiva, astratta o computazionale, poiché il sapere genera effetti che investono la percezione, la comprensione,

27 - Questo approccio si avvicina a prospettive fenomenologiche, come quelle di Edmund Husserl e Maurice Merleau-Ponty. Husserl, nella sua analisi della crisi delle scienze europee, insiste sul legame tra conoscenza e esperienza vissuta (*Erlebnis*), mentre Merleau-Ponty, nella *Phénoménologie de la perception*, mostra come la percezione e il sapere siano sempre incarnati e situati. In ambito epistemologico, Michael Polanyi ha parlato di una "conoscenza personale", sottolineando che il sapere autentico non è mai puramente oggettivo, ma implica un coinvolgimento tacito del soggetto. Cfr. E. Husserl, *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*, Martinus Nijhoff, Haag 1954; Maurice Merleau-Ponty, *Phénoménologie de la perception*, Gallimard, Paris 1945; Michael Polanyi, *Personal Knowledge: Towards a Post-Critical Philosophy*, University of Chicago Press, Chicago 1958.

il valore e l'identità epistemica di chi conosce. In breve, l'IA produce risultati simbolici utilizzabili, ma non attraversa il sapere: non ne fa esperienza, e non è trasformata da esso. Se la transdisciplinarità è un sapere che modifica il soggetto conoscente, l'IA non può praticarla, perché non è un soggetto. Può rielaborare informazioni, generare risposte sofisticate, persino simulare processi creativi, ma non partecipa realmente al sapere. Il suo contributo rimane quindi confinato a una dimensione puramente operativa, senza entrare nel campo della trasformazione epistemica che caratterizza il sapere umano. Questa differenza segna un confine netto tra l'IA e la transdisciplinarità: mentre l'IA eccelle nella gestione di dati e nella creazione di modelli predittivi, la costruzione di significati e la trasformazione del soggetto restano prerogative dell'intelligenza umana. Il sapere transdisciplinare, quindi, non può essere delegato all'IA, perché richiede un coinvolgimento che va oltre il calcolo e tocca la natura stessa del conoscere.

5. Sul Ruolo delle Università nell'Era dell'Intelligenza Artificiale

Siamo giunti a un passaggio decisivo. Ci troviamo di fronte a una svolta epistemica e formativa che interpella l'essere umano non solo nelle sue competenze, ma nella sua stessa vocazione a conoscere e ad abitare il mondo in modo significativo. L'intelligenza artificiale non è più solo uno strumento, ma un attore autonomo nella produzione di conoscenza tecnica e operativa, capace di superare le capacità umane in molti ambiti analitici e computazionali. Se il sapere tecnico è sempre più assorbito dall'IA, il rischio è che il ruolo dell'essere umano nella conoscenza diventi residuale, limitato a supervisionare e validare processi che ormai si sviluppano autonomamente. Eppure, nella sezione precedente abbiamo argomentato che esiste uno spazio essenziale per l'intelligenza umana, e questo spazio è la transdisciplinarità. La capacità di tenere insieme la complessità, di costruire connessioni tra saperi diversi, di generare nuove prospettive e di superare la frammentazione della conoscenza è qualcosa che l'IA, per come è strutturata, non può praticare. Ma se la transdisciplinarità è ciò che può restituire all'essere umano un ruolo centrale nella produzione di conoscenza, le nostre istituzioni educative e scientifiche sono pronte a riconoscerne la centralità e a integrarla nei loro modelli epistemici e formativi?

Qui emerge un problema. Oggi è il sapere tecnico a dominare, non solo nel mondo del lavoro, ma nelle università stesse. I corsi di laurea sono progettati per formare specialisti con competenze iper-focalizzate, mentre pensiero laterale, capacità di sintesi, ragionamento analogico e unità del sapere sono considerati marginali. La transdisciplinarità, più che un obiettivo formativo, sembra essere un lusso intellettuale per pochi, mentre la specializzazione viene percepita come il requi-

sito fondamentale per avere successo. Ma questa impostazione ha ancora senso nell'era dell'IA? Se il sapere tecnico sta diventando dominio delle macchine, ha senso che il sistema educativo umano continui a concentrarsi solo su quello? O forse le università dovrebbero ripensare la loro missione, non più come semplici fabbriche di specialisti, ma come luoghi in cui il sapere viene generato, integrato e trasformato? Nel libro *Università Generativa*, Andrea Prencipe analizza proprio questa sfida: il passaggio da un'università che trasmette nozioni a un'università che promuove forme di pensiero capaci di attraversare la complessità28. Se vogliamo che l'essere umano mantenga un ruolo centrale nella costruzione della conoscenza, allora le università devono essere il luogo in cui questa capacità viene coltivata. La questione è chiara: se il sapere tecnico, in un certo senso, sarà sempre più affidato all'IA, quindi con il ruolo giocato dagli esseri umani sempre più marginale, il compito delle università non può essere quello di formare semplicemente specialisti, ma deve diventare quello di preparare persone capaci di dare direzione e significato alla conoscenza stessa. La transdisciplinarità non può più essere un aspetto secondario della formazione: deve diventare il fondamento di un nuovo paradigma educativo, in cui la comprensione profonda del mondo non è più il sottoprodotto della specializzazione, ma il cuore stesso del sapere. In questa sezione intendo proporre alcuni elementi utili ad avviare una riflessione su come le università possano accompagnare questo tempo di transizione, diventando luoghi di cambiamento, rigenerazione e produzione di sapere generativo. Non si tratta di una trattazione esaustiva, ma di un insieme di spunti pensati per aprire a una riflessione più ampia.

Un primo aspetto fondamentale è quello formativo e pedagogico. Se il sapere tecnico è destinato a essere sempre più assorbito dall'IA, le università devono interrogarsi sul tipo di conoscenza che intendono trasmettere. Oggi, il modello accademico privilegia la formazione di specialisti disciplinari, esperti capaci di padroneggiare in modo approfondito un ambito specifico, ma spesso senza strumenti per integrare e attraversare i confini tra discipline. Questo paradigma, già oggi insufficiente a comprendere la complessità del reale, rischia di diventare obsoleto in un contesto in cui le competenze strettamente tecniche saranno sempre più delegate all'intelligenza artificiale. Le università devono quindi diventare spazi di formazione transdisciplinare, capaci di preparare non solo specialisti, ma persone in grado di operare su più piani del sapere. A livello interdisciplinare, ciò implica la capacità di connettere conoscenze e metodi provenienti da discipline differenti per affrontare problemi complessi in modo cooperativo e integrato. A un livello più profondo, transdisciplinare, significa attraversare i confini disciplinari per accede-

28 - A. Prencipe, *Università Generativa*, il Mulino, Bologna 2024.

re a diversi livelli di realtà, integrando prospettive epistemiche, etiche e simboliche che non si lasciano ridurre a una sintesi puramente metodologica. Le "visioni integrate" non sono semplici compendi di saperi, ma orizzonti generativi che emergono dalla capacità di riconoscere le interazioni tra sistemi differenti, di abitare l'ambiguità e di generare significato in contesti ad alta complessità. Un'educazione di questo tipo non si limita a trasmettere informazioni, ma allena alla flessibilità cognitiva, alla capacità di superare dicotomie rigide e di navigare tra domini ontologici distinti. Se vogliamo che l'essere umano mantenga un ruolo centrale nella produzione del sapere, le università devono farsi carico di questa transizione: non più soltanto luoghi di addestramento tecnico-specialistico, ma ambienti generativi di pensiero trasformativo.

Oltre alla dimensione pedagogica, le università dovrebbero diventare anche spazi di ricerca transdisciplinare, in cui scienza, filosofia, teologia, arte e altri saperi non si limitino a convivere, ma collaborino nell'elaborazione di orizzonti condivisi. In questa prospettiva, la filosofia non è né un sapere dominante né un'ancella degli altri, ma un interlocutore critico essenziale, capace di interrogare i fondamenti, aprire domande radicali e accompagnare l'articolazione di una comprensione plurale e situata della realtà. L'obiettivo non è una "riflessione unitaria" intesa come sintesi sistematica, ma una co-costruzione di senso che attraversi i diversi livelli del reale e li metta in relazione, mantenendone la specificità epistemica. In questo processo, la filosofia assume il ruolo di catalizzatore e garante della coerenza riflessiva, senza però sovrapporsi o assorbire gli altri linguaggi del sapere. Oggi, tuttavia, questa prospettiva è ostacolata dall'attuale organizzazione accademica, fortemente segmentata in dipartimenti e discipline separate. Il modello universitario è costruito sulla logica della specializzazione e della divisione del sapere: le discipline dialogano, ma restano confinate nei loro ambiti, con interazioni spesso limitate a collaborazioni occasionali tra specialisti. Ma la transdisciplinarità, intesa nel senso più profondo, non è una semplice collaborazione tra esperti di settori diversi, né la somma di competenze disciplinari che ogni tanto si incontrano per lavorare su un problema comune. Il metodo transdisciplinare, di cui qui non siamo riusciti a trattare in dettaglio e che richiederebbe una riflessione più ampia, non si basa su una semplice integrazione di discipline esistenti, ma implica l'emergere di una soggettività conoscente rinnovata – individuale o collettiva – capace di attraversare livelli di realtà differenti, di scoprire connessioni inaspettate e di generare nuovi paradigmi. Si tratta di un soggetto trasformato dall'incontro con la complessità, che non osserva il mondo da una posizione esterna, ma vi partecipa attivamente, modificandolo e modificandosi nel processo conoscitivo. Affinché questo avvenga, serve un ripensamento non solo metodologico, ma strutturale. La suddivisione attuale in dipartimenti e facoltà è funzionale a una ricerca specialistica, ma

forse non è il modello più adatto per chi si propone di attraversare le discipline e costruire un sapere integrato. Servono spazi, tempi e modi che favoriscano l'emergere di una conoscenza transdisciplinare. Questo potrebbe significare l'istituzione di centri di ricerca transdisciplinari all'interno dell'università, o addirittura il superamento della suddivisione rigida tra dipartimenti. Ma più ancora, occorre promuovere comunità epistemiche capaci di coabitare la complessità: gruppi di ricerca in cui soggetti provenienti da ambiti diversi si confrontano in modo dialogico, trasformandosi a vicenda nel processo stesso del conoscere. La transdisciplinarità, infatti, non è soltanto un metodo, ma una pratica comunitaria in cui il sapere emerge dalla relazione – tra persone, tra livelli di realtà, tra visioni del mondo.

Costruire queste comunità significa riconoscere che la conoscenza non si produce in solitudine, ma nasce dall'ascolto, dal conflitto fecondo e dalla fiducia reciproca che rendono possibile l'emergere di paradigmi condivisi. Questa è una sfida ancora aperta. Se vogliamo che le università non si limitino a trasmettere e specializzare, ma diventino luoghi di generazione di un sapere realmente nuovo, sarà necessario ripensarne le strutture e i modelli di ricerca. Forse il futuro della conoscenza non sta semplicemente nel migliorare le interazioni tra discipline, ma nel creare un contesto in cui il sapere possa fluire liberamente tra di esse, favorendo la nascita di nuovi soggetti epistemici plurali – individui e comunità capaci di attraversare la complessità del reale, generando visioni trasformative.

Un ultimo punto, forse il più profondo, riguarda il ruolo culturale delle università. Il compito dell'università non è solo quello di formare specialisti o di produrre ricerca avanzata, ma anche di traghettare l'intera umanità in una nuova era, in cui l'essere umano sia al centro per quello che è il suo specifico.

Viviamo in un tempo in cui il sapere tecnico cresce esponenzialmente, ma senza un orizzonte di senso rischia di diventare frammentato e privo di direzione. L'università non può limitarsi a trasmettere competenze tecniche: deve mediare tra sapere scientifico e sapere umanistico, tra innovazione tecnologica e riflessione sul significato della conoscenza stessa. Non basta saper fare: occorre anche comprendere il significato di ciò che si fa, e il contesto umano e storico in cui lo si fa. Questo significa che l'università non è solo un luogo di formazione e ricerca, ma un luogo di narrazione e di sintesi, uno spazio in cui le connessioni tra saperi vengono esplicitate e trasformate in una visione unitaria.

La divulgazione, in questo senso, non è semplicemente la trasmissione di informazioni tecniche, ma il racconto di una storia più ampia, capace di dare senso al sapere e di ricollocarlo in un quadro comprensibile e condiviso. Per superare l'ultratecnicismo e restituire all'essere umano il suo ruolo nella conoscenza, dobbiamo raccontare storie che dicano l'unità del sapere, che mostrino connessioni profonde tra discipline, che offrano un quadro in cui ogni individuo possa ricono-

scersi, sognare, comprendersi. Questa dimensione narrativa non è una riduzione retorica della transdisciplinarità, ma ne è al contrario uno degli esiti più alti: il racconto consente di integrare piani diversi dell'esperienza – scientifico, etico, estetico, esistenziale – e di restituire al sapere la sua radice incarnata, storica, condivisa. La transdisciplinarità non produce solo conoscenze astratte, ma orizzonti di significato, e questi orizzonti si articolano anche attraverso il linguaggio narrativo. Narrare, in questo senso, significa ricomporre la frammentazione del sapere, non semplicemente decorarla.

Senza questo orizzonte, il sapere rischia di disperdersi in un'infinità di dettagli, senza mai riuscire a costruire una visione del mondo. Se l'università vuole essere ancora il cuore della conoscenza nell'era dell'IA, deve farsi carico di questo compito: non solo formare specialisti, ma offrire un orizzonte di senso che possa guidare la trasformazione del sapere e dell'umanità stessa.

6. Conclusioni

L'intelligenza artificiale sta trasformando radicalmente il panorama della conoscenza, ridefinendo il ruolo dell'essere umano nella produzione del sapere. Se il sapere tecnico e operativo è sempre più assorbito dall'IA, ciò non significa che l'essere umano sia destinato a diventare marginale. Abbiamo cercato di mostrare che esiste una dimensione della conoscenza che resta essenzialmente umana: la transdisciplinarità. L'IA eccelle nell'analisi dei dati, nella correlazione e nella previsione, ma non genera nuovi livelli di realtà, non naviga la complessità con il principio del terzo incluso, non ha un rapporto soggetto-oggetto che le permetta di trasformare sé stessa attraverso la conoscenza. Per questo, la transdisciplinarità non è solo un'alternativa, ma una necessità: è il sapere che permette di integrare le specializzazioni, di costruire significati e di generare nuove prospettive.

Le università, in questo scenario, hanno un compito decisivo. Devono essere il luogo in cui questa capacità viene coltivata, non solo attraverso una formazione transdisciplinare, ma anche ripensando le strutture della ricerca e restituendo un orizzonte di senso alla conoscenza. Se il sapere frammentato rischia di perdere direzione, l'università deve diventare il luogo in cui le connessioni tra i saperi vengono ricostruite e narrate. In definitiva, la sfida non è semplicemente quella di integrare l'intelligenza artificiale nei processi conoscitivi, ma di preservare e rilanciare ciò che rende la conoscenza autenticamente umana: la capacità di significare, di trasformarsi nel conoscere, di attraversare livelli di realtà e generare visioni condivise.

In questa prospettiva, l'IA non è un avversario né un sostituto, ma un banco di prova per ridefinire con più consapevolezza la responsabilità dell'essere umano nel custodire e sviluppare il senso del sapere. Il futuro della conoscenza non sarà

deciso solo dall'efficienza degli algoritmi, ma dalla nostra capacità di coltivare spazi di pensiero capaci di unificare, trasformare e orientare.

MARCO SANCHIONI
Professore straordinario di Filosofia della scienza presso l'Istituto Universitario Sophia (FI)
marco.sanchioni@sophiauniversity.org

Vulnerabilità creativa e intelligenza artificiale

di Antonio Bergamo

The contribution explores the relationship between artificial intelligence and the human condition, questioning how AI challenges contemporary anthropology. It introduces the notion of creative vulnerability as a key to understanding what is unique to the human being in contrast to the increasingly sophisticated capabilities of artificial systems. The essay thus investigates the in-between – the relational and intersubjective space – in which the human subject is constituted through the capacity to feel, create, fail, and respond to otherness. It ultimately invites a rethinking of thinking and creating as ethical and existential practices, in which vulnerability is not merely a limitation, but also a generative resource.

Keywords: Vulnerability; Creativity; Relationality; Otherness; Intelligence

Lo sviluppo dell'intelligenza artificiale costituisce una interpellanza per l'antropologia[1]. Nel mentre l'IA sta compiendo un progresso significativo, passo dopo passo, noi come esseri umani siamo oscillanti tra lo stupore e la preoccupazione. Ma non è solo questa reazione a determinare un interrogativo da cui approfondire la questione. Risulta ancora più decisivo domandarsi in che modo essa stia plasmando il nostro modo di abitare il mondo e le relazioni[2].

Ogni relazione porta in sé una trasformazione, e tale trasformazione non è per noi un mero automatismo della carne che reagisce ad una interazione, ma coinvolge in maniera più o meno consapevole i vari livelli della complessità che ci segna. Quando ci rapportiamo a sistemi artificiali di intelligenza ci pare di poter dialogare in maniera convincente con loro, rimaniamo stupiti dalle loro produzioni, in qualche modo ci rassicura che riescano a diagnosticare una malattia talvolta meglio degli esseri umani, siamo però anche angosciati di come essi possano in qualche modo sostituirci[3].

La questione tuttavia non è sino a che punto un sistema artificiale di intelligenza possa raggiungere o superare l'intelligenza umana, ma in che modo gli umani possano fare una esperienza che sia arricchita dal supporto dell'intelligenza artificiale e non ne sia invece impoverita, rinnovando il dramma di Dedalo che in questo caso vedrebbe il precipitare anzitutto nel perdere per strada quelle risorse generative tipiche dell'umano. Una di queste risorse generative può essere considerata la creatività e la sua alleata, la vulnerabilità. È in tal senso che si può delineare una nozione di vulnerabilità creativa quale "tra" della relazionalità reciproca e sistemica che intesse il reale e gli umani.

---

1 - Per una chiarificazione terminologica si assuma come punto di partenza la definizione di Francesca Rossi: «l'intelligenza artificiale è una disciplina che mira a sviluppare programmi o macchine (software e/o hardware) che mostrano un comportamento che verrebbe definito intelligente se fosse esibito da un essere umano» (F. Rossi, *Il confine del futuro. Possiamo fidarci dell'intelligenza artificiale?*, Feltrinelli, Milano 2019, p. 25). Daniel Andler distingue inoltre tra l'intelligenza artificiale in quanto oggetto che si cerca di creare e come disciplina in un senso più ampio che cerca di caratterizzare e costruire tale oggetto. Emerge inoltre la nozione di sistema artificiale intelligente per indicare un sistema prodotto dall'intelligenza artificiale e che, nel linguaggio comune, viene definito *una* intelligenza artificiale: cfr. D. Andler, *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, Einaudi, Torino 2024, pp. 5-9.

2 - A tale riguardo Claverini ha inteso evidenziare l'impattare nella porosità dell'esistenza individuale dell'intelligenza artificiale: C. Claverini, *Etica sintetica e nuove identità artificiali*, in «Segni e Comprensione», 108/2024, 102-115.

3 - Cfr. M. Dorobantu, *Interpréter l'humain – L'* Imago Dei *à l'heure du numérique*, in «Recherche de Science Religieuse», 4/2023, pp. 661-678.

1. L'umano tra pensiero computazionale e discernimento del senso dell'essere

In questo quadro la creatività rappresenta uno snodo decisivo per delineare una postura antropologica significativa. La creatività è una prerogativa solo umana? Se un sistema artificiale di intelligenza produce una immagine, un testo, una musica di buona se non ottima qualità può essere considerato creativo? Certamente la creatività è un processo complesso e profondo[4]. Esso coinvolge immaginazione, ispirazione e innovazione. L'intelligenza artificiale, si può già affermare, esegue compiti ripetitivi, ma partecipa anche attivamente alla produzione artistica, musicale, letteraria.

Markus Gabriel ha giustamente notato come l'IA si rapporti all'intelligenza umana come la mappa al territorio[5]. Essa si offre a noi non come una questione di pensiero, ma come modelli di pensiero. L'IA pertanto è un processo di modellizzazione, in cui un modello deve al meglio somigliare a ciò di cui è modello. Non è una mera copia, tuttavia, ma può avere altre proprietà rispetto a quelle che noi possiamo comprendere o spiegare attraverso di essa. L'essere umano che è un soggetto in relazione non rischierebbe in tal senso di pensarsi come un momento del processo di modellizzazione lì dove perdesse il baricentro della sua esperienza? Se esperienza è anche ciò che mi accade, allora essa non si riduce a una mera modellizzazione, come suo primo e originario momento, ma come una co-creazione di significato in un campo di senso intersoggettivo.

È noto che l'informatica come noi la conosciamo oggi sia un prodotto della logica[6]. Se Einstein rammentava come «la logica vi farà andare da A a B. L'immaginazione vi farà andare ovunque», ed ancora, «l'immaginazione è più importante della conoscenza. Perché la conoscenza è limitata a quello che conosciamo e comprendiamo, mentre l'immaginazione abbraccia il mondo intero, e tutto quello che mai ci sarà da conoscere e capire»[7], è perché, come evidenzia Daniel Andler, gio-

4 - Lo psicologo Philip Johnson-Laird ha proposto una teoria computazionale del processo creativo, chiedendosi se la creatività è simulabile su un dispositivo digitale, ed evidenziando il carattere non deterministico della creatività: P.N. Johnson-Laird, *Deduzione induzione creatività. Pensiero umano e pensiero meccanico*, Il Mulino, Bologna 1994; M.R. Stufano Melone, *Ontologie della creatività*, Franco Angeli, Milano 2019, pp. 70-73. Teresa Amabile ha elaborato a sua volta un modello componenziale della creatività, in cui accanto alle competenze, ai processi, alla motivazione al compito, svolge una significativa funzione l'ambiente sociale: T.M. Amabile, *Creativity in Context: Update to the Social Psychology of Creativity*, Taylor & Francis Limited, New York 2019.

5 - Cfr. M. Gabriel, *Pourquoi la pensée humaine est inégalable*, JC Lattès, Paris 2019, pp. 114 ss.

6 - Cfr. D. Andler, *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, cit., pp. 36-61.

7 - A. Einstein, *The Ultimate Quotable Einstein*, ed. by A. Calaprice, Princeton University Press, Princeton 2011, pp. 481 e 12.

cano un ruolo decisivo negli umani i sentimenti epistemici, i quali contribuiscono al processo di conoscenza e alla soluzione dei problemi[8].

I sistemi artificiali di intelligenza sono una logica pura che noi implementiamo su *hardware* ed obbediscono alle leggi della fisica che ci sono note, le quali si possono esprimere in equazioni matematiche. Se, come sostiene Luciano Floridi, l'intelligenza è la facoltà di pensare e l'intelligenza artificiale ha introdotto un paradigma più computazionale e informazionale nella riflessione filosofica[9], sarebbe necessario chiedersi cosa significa pensare e, ancor più radicalmente, come ha fatto Heidegger, che cosa ci chiama a pensare[10]. La nostra facoltà di pensare si confronta quotidianamente con accadimenti previsti ed imprevisti che riconfigurano e ricompaginano il nostro stare al mondo. L'esperienza pertanto non è solo ciò che ci accade, ma anche ciò che attraversiamo. Ed alcuni accadimenti non sono sempre intenzionali. Anzi, molto spesso la nostra vita sia a livello micro che macrosistemico dice uno sviluppo che contiene fatti che non sono stati pianificati da qualcuno. Lo stesso Floridi sembra essere consapevole di questo: noi, afferma il filosofo, reagiamo al non senso della vita, alla sua mancanza di significato, fabbricando una infosfera, la quale diventa in pratica la nostra atmosfera spirituale[11]. Ma il senso della nostra vita spirituale si esaurisce in un processo informazionale? L'IA con la sua potenza generativa giocherebbe un ruolo non da poco in questo sentimento dell'angoscia del non senso: da un lato essa ci turba portando l'interrogativo sul nostro essere noi stessi e dall'altro ci rassicura, promettendo di togliere lo stress di stare al mondo, proiettando in essa il nostro horror vacui, nella speranza che lo colmi.

Eppure la nostra esperienza non si riduce al mero dato. Non c'è infatti coscienza intenzionale senza una interazione con la coscienza fenomenale[12]. La mappa non è il territorio, per cui il territorio nel suo essere in divenire, produce una sempre nuova interpellanza negli umani. L'intelligenza umana è in tal modo fondamentalmente emozionale. L'esperienza è granulare, costituita da fini elementi che la attraversano, ed è simultaneamente individuale ed intersoggettiva. Quando ad

8 - Cfr. D. Andler, *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, cit., pp. 246-248.

9 - Cfr. L. Floridi, *Filosofia dell'informazione*, Raffaello Cortina, Milano 2024, pp. 14-15.

10 - Cfr. M. Heidegger, *Che cosa significa pensare?,* Sugarco, Milano 1996, pp. 153-154.

11 - «La vita mentale è quindi il risultato di una reazione riuscita a un originario *horror vacui* semantico: il caos privo di significato (nel senso non esistenzialista di "non ancora significante") minaccia di fare a pezzi il Sé, di farlo sprofondare in un'alterità alienante percepita dal Sé come il nulla; questo timore primordiale dell'annichilimento spinge il Sé a riempire incessantemente ogni spazio semanticamente vuoto con qualsiasi significato che il Sé riesca a mettere insieme, con il massimo successo consentito dal contesto dei vincoli, delle possibilità e dello sviluppo della cultura» (L. Floridi, *Filosofia dell'informazione*, cit., p. 20).

12 - Cfr. M. Gabriel, *Pourquoi la pensée humaine est inégalable*, cit., pp. 141-145.

esempio mi imbatto con la morte di altri, sono interrogato dal fatto che la sottrazione dell'altro al mio campo visivo, non sia semplicemente una sottrazione di significato, ma il non esserci più dell'altro invoca un nuovo significato a quella esperienza di assenza e di vuoto, per cui dal dolore del distacco sono portato a interrogarmi sul mio nuovo essere là insieme con altri nel percorso della vita. L'intelligenza umana che pure contiene in sé una dimensione computazionale, non si esaurisce in essa. Essa è più delle sue rappresentazioni. Nemmeno l'intelligenza è solo un insieme di funzioni tese ad un obiettivo. In tale prospettiva il materialismo, il naturalismo e il funzionalismo che, pur con le loro differenze, si fondano sull'assunto che mente e cervello umani siano sostanzialmente identificabili e funzionino come i loro modelli computazionali, non sono ipotesi scientifiche dimostrabili, bensì interpretazioni metafisiche della realtà, accomunate da un «riduzionismo computazionale» teso a riformare l'ontologia umana[13]. Per quanto alcuni autori come Dennett critichino la religione e la metafisica come non scientifiche, anche materialismo, naturalismo e funzionalismo si appoggiano ad un modello di pensiero[14].

Tutta l'esperienza umana è attraversata da cima a fondo da una relazionalità nella quale la logica emerge come il darsi di reciproche interazioni da cui è possibile attingere un significato potenziale. E la relazionalità umana è fondata sulle nostre imperfezioni e limiti cognitivi[15].

L'IA si presenta come una forma di intelligenza computazionale, capace anche di emulare gli stati emozionali e di riconoscere in qualche modo quelli dei suoi interlocutori umani. Nel campo della psicologia Lisa Feldman Barret, tuttavia, ha elaborato la sua teoria delle emozioni costruite, in cui le emozioni non sono mere risposte a qualcosa che sopraggiunge da fuori[16]. Piuttosto la complessità della soggettività umana è coinvolta in un intreccio micro e macrosistemico, sia fisiologico che culturale, nel quale i significati si originano a partire da un articolarsi di risonanze affettive che hanno struttura fisiologica e slancio cognitivo. Il cervello non si ridurrebbe quindi solo alla mente e il pensiero non sarebbe semplicemente ciò che è in funzione di qualcosa, ma una tessitura strettamente legata ad un sostrato di ordine biologico, sorpassandolo al contempo. I significati si vanno costruendo nel mentre noi procediamo nell'esperienza della nostra vita, che è strettamente legata a quella degli altri. L'intelligenza artificiale è invece «disincarnata», «non avendo

---

13 - F. Pianca, *La narrativa dell'intelligenza artificiale: soggettività etica e riduzionismo antropologico*, in «Anthropologica», 2020, p. 88.

14 - Cfr. M. Gabriel, *Pourquoi la pensée humaine est inégalable*, cit., p. 153.

15 - Cfr. M. Dorobantu, *Interpréter l'humain – L*'Imago Dei *à l'heure du numérique*, cit., p. 677.

16 - Cfr. L. Feldman Barrett, *Come sono fatte le emozioni. La vita segreta del cervello*, Giunti, Firenze 2023.

corpo, essa non ha un sé»[17]. Un sistema artificiale di intelligenza può modellizzare dunque un processo di pensiero e simularlo, ma il punto di partenza e di approdo non risiederebbe nel "tra" di una esperienza incarnata, quanto nel "tra" computazionale, che sarebbe uno solo dei livelli cognitivi.

2. Creatività e pensiero

Si tratta allora di chiedersi che cosa intendiamo per creatività. Solitamente nel senso comune affermiamo che qualcuno è creativo quando inventa o scopre qualcosa, quando mette al mondo una novità. Il matematico Jacques Salomon Hadamard distingue tra invenzione, riferita a qualcosa di già non esistente; e scoperta, riferita invece a qualcosa o qualche proprietà di qualcosa che già esiste[18]. Di per sé questo lo conduce a ipotizzare che per l'artista vi sia più libertà di invenzione, mentre per lo scienziato prevarrebbe l'elemento della scoperta. Tuttavia, annota Gabriele Lolli, «in campo artistico non si direbbe mai che s'inventa un'opera, come un quadro, una scultura. Non inventiamo un'opera d'arte, ma neanche la scopriamo, la parola più adatta sarebbe "creare", che per la maggior parte delle persone è più affine all'invenzione che alla scoperta»[19].

Nella "creazione" allora si inventa o si introduce magari uno stile, come può essere ad esempio un genere letterario. L'IA componendo e collegando una miriade di dati è capace di simulare uno stile e di ricalcarlo, forse può o potrà produrre un nuovo stile, ma sarebbe necessario chiedersi cosa sia uno stile e quale sia il regime epistemico che lo segna. Resta però che il punto di partenza dell'IA nella produzione di un qualcosa riposi in una dinamica combinatoria che le è controllabile su segmenti specifici, mentre nell'intelligenza umana la creazione di qualcosa non è controllabile dal soggetto umano. Così un sistema artificiale di intelligenza può essere un formidabile giocatore di Go ma un completo disastro a scacchi, perché il suo apprendimento si base su un campo specifico e il suo apprendimento si allena sui dati relativi a quel campo[20].

17 - D. Andler, *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, cit., p. 241.

18 - Cfr. J. Hadamard, *La psicologia dell'invenzione in campo matematico,* Raffaello Cortina, Milano 1993.

19 - G. Lolli, *La creatività in matematica*, Bollati Boringhieri, Torino 2024, p. 13.

20 - In particolare nel *deep learning* emerge con i modelli dell'IA traggono la loro efficacia dal prendere in considerazione casi particolari, i dati, in una quantità massiccia: cfr. D. Andler, *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, cit., pp. 248-250.

Se, come rileva Lolli, le scoperte sono oggetti o fatti sui quali non si ha un controllo, e le invenzioni e le creazioni, sono oggetti o procedure con caratteri che possono essere scelti dall'inventore o dal creatore, il discorso si complica[21].

Il procedimento che porta l'IA ad una possibile scoperta è sempre governato da una cablatura, per così dire, che le è stata implementata, mentre per l'intelligenza umana questo procedimento sembra essere sempre legato a ciò che sentiamo, percepiamo e facciamo, a sua volta legato con ciò che gli altri sentono, percepiscono, fanno o dicono. Questo fattore implica un duplice riscontro: che la creazione sia co-creazione, anche quando portata avanti nella solitudine di un lavoro personale, in quanto condivisa e realizzata con altri in un determinato luogo o situazione, che è l'atmosfera spirituale nella quale i vissuti si intrecciano; che la creazione porti in sé il marchio della vulnerabilità, in quanto sono esposto alla risonanza affettiva che mi interpella dinanzi al vissuto dell'altro, al peso del mio stare al mondo, all'errore che inevitabilmente posso commettere.

Un sistema artificiale di intelligenza non ha situazioni nelle quali è immerso ed implicato, il suo mondo «è popolato solo da simboli»[22], a differenza dell'essere umano che è immerso nelle situazioni, le quali producono per lui una responsabilità, ovvero una capacità di risposta. Si rivela preziosa pertanto l'indicazione di José Ortega y Gasset quando nelle *Meditazioni del Chisciotte* considera che «io sono io e le mie circostanze»[23]. Ogni atto di coscienza non implode né si esaurisce in una qualsiasi riflessione. Per cui ciò che io percepisco non è identico a se stesso nell'atto della riflessione e per riuscire invece a pensarlo autenticamente dovrò provare a comprenderlo nella sua esecutività. Quando ascolto Mozart qualcosa accade: vi è un passaggio tra la musica di Mozart e me, tra udire la musica di Mozart e sentire di udire la sua musica. Quello che accade in questo passaggio è delimitato dalle bordature della polifonia del mondo, per cui quella musica è di Mozart ma è anche la musica di Mozart in me, e quindi il mio suono. Quando io poi eseguo Mozart allora la mia interpretazione stessa di Mozart è *altro*, e si attiva un processo creativo che non posso prevedere in partenza dove mi condurrà. Il tentativo di Ortega y Gasset si muove nel superamento sia di un realismo ingenuo che di un idealismo asfittico e la sua ragione vitale focalizza il pensiero come l'interpretare la vita che si dà nell'atto stesso di viverla. La riflessione di Ortega verrà ripresa, integrata e superata da María Zambrano, per cui la ragione vitale è ragione poetica o, per meglio dire, ragione poietica e creativa in cui la compaginazione di *logos* e *pathos* accade in una differente *claritas*, chiarezza cognitivo-emotiva, intesa quindi come regione

21 - Cfr. G. Lolli, *La creatività in matematica*, cit., pp. 15-16.

22 - D. Andler, *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, cit., p. 241.

23 - Cfr. J. Ortega y Gasset, *Meditazioni sulla felicità*, SugarCo Edizioni, Milano 1994.

formale, campo di senso formale, in cui in un processo che è individuale ed intersoggettivo al contempo, la vita si va facendo, ovvero lascia emergere una novità creativa[24].

3. La creatività del sentire

Cosa è dunque "creazione"? Essa viene in rilievo come la creatività del sentire, percezione che non si riduce alle sue rappresentazioni[25], portando con sé la capacità di suscitare lo stupore. Questa creatività non nascerebbe dal nulla, ma poggia nella complessità di fili che intessono il reale in cui perviene all'esistenza qualcosa che non può essere circoscritto né alla volontà di qualcuno né a fatti meramente positivi[26].

Il campo semantico della creatività porta in sé tale intermezzo. Ad esso possono essere ricondotti i termini *creazione*, *creatore* e *creatura*. Bernhard Waldenfels le ha definite *parole concettuali nomadi*, per cui non trovano mai una collocazione univoca e sfuggono al contempo ad una mera equivocità.

La parola *creazione* fa volgere subito il pensiero alla mitologia, alla teologia e alla creazione del mondo. Al creatore del mondo poi non si è soliti collegare il termine creatività e «proprio perché non sarebbe creativo un Creatore (del mondo) o Sommo artefice, è considerato una creazione incarnata e una incarnazione della creatività o creatività per essenza»[27].

Il termine *creatività* ha poi fatto la sua comparsa con singolare pregnanza nella letteratura scientifica della psicologia a partire dagli anni trenta del secolo scor-

---

24 - Cfr. M. Zambrano, *Chiari del bosco,* Bruno Mondadori, Milano 2004.

25 - In questo senso G. Cusinato propone di pensare «la sensazione (*Empfindung*) e il sentire (*Fühlen*) come due modalità fondamentalmente diverse che la vita ha a disposizione per rapportarsi alla dimensione espressiva: ambedue si avvalgono di organi sensibili, ma mentre la sensazione organica coglie l'espressività in funzione della semiosi che governa l'interazione fra organismo e mondo ambiente (*Umwelt*), il sentire è meta espressivo, nel senso che coglie l'espressività di qualcosa in funzione dello sviluppo e riuscitezza di un percorso espressivo di formazione. Mentre il sentire si rivolge fondamentalmente ai valori del bello (fecondità) e del brutto (sterilità), la sensazione coglie in primo luogo quelli del piacevole e del doloroso. In tal modo il sentire reinterpreta l'esperienza secondo una propria sfera di rilevanza e, in alcuni casi, dimostra di avere un proprio accesso, diretto e indipendente, alla dimensione espressiva dell'esperienza, come quando coglie nel rossore del volto di una persona il pudore o la vergogna, senza aver bisogno di dedurli dal contenuto percepito della mera sensazione organica del rosso» (G. Cusinato, *Periagoge. Teoria della singolarità e filosofia come esercizio di trasformazione,* QuiEdit, Verona 2017, p. 89).

26 - Cfr. R. Guccinelli, *"Nelle crepe e ai margini del quotidiano": la creatività responsiva di Bernhard Waldenfels*, in B. Waldenfels, *Creatività responsiva*, Inschibboleth, Roma 2022, p. 17.

27 - B. Waldenfels, *Creatività responsiva*, cit., pp. 111.

so. In particolare nella psicologia della *Gestalt,* in cui però si tende a parlare di più di produttività, di pensiero produttivo, distinto da quello riproduttivo, il quale sarebbe incapace di ristrutturare in maniera originale il campo cognitivo o percettivo.

Creazione assume poi differenti connotazioni nei contesti linguistici in cui compare. Il termine latino *creatio* porta con sé una sottolineatura del generare e del crescere, legato in qualche modo al termine greco *poiesis,* pensato invece sul modello del produrre artigianale, mentre il termine ebraico *barah*, presente nel racconto della creazione, è connesso alla creazione del mondo concepita come una creazione senza strumenti di un qualcosa di nuovo. Sono quindi intonazioni semantiche che presentano una topografia variegata.

Il pensare porta in sé custodito non il darsi di una conoscenza sempre chiara, ma l'albeggiare in essa di un senso possibile e di una risposta. Il *logos*, l'interpretazione che noi diamo della realtà, è combinato con la situazione concreta in cui noi siamo immersi, che è fatta di carne e sangue, *pathos*, di una storia che ci precede, in cui siamo immersi e che contribuiamo a scrivere. Motivo per cui non viviamo sempre al presente, e la creazione di qualcosa sembrerebbe essere l'irruzione in un istante di qualcosa che poi va integrandosi nel campo di senso della nostra situazione e delle situazioni degli altri.

Ogni accadimento domanda una risposta, e la nostra risposta non è mai solo la nostra reazione ad esso, ma un processo creativo in cui il significato viene ad essere costruito. La chiarezza cognitivo-emotiva che si dà nella penombra, intuita da Zambrano, ancora una volta si rivela come la regione formale in cui qualcosa si va producendo *tra* accadimento e risposta. Quando qualcosa mi viene in mente, quando si affaccia una nuova intuizione o un pensiero, questo non è mai un mero momento soggettivo, ma si rivela in un intreccio, per cui gli umani sono «soggetti esperienti»[28]. È, quella descritta da Zambrano, la fase della costruzione dei significati o formazione delle regole, in maniera tale, come evidenzia dal suo punto di vista Waldenfels, che «ciò da cui veniamo affezionati se non ci fossero eventi lampo e interpretazioni si trasforma in qualcosa che viene appercepito, trattato, desiderato ardentemente, compreso in quanto qualcosa»[29].

4. Il darsi di una novità

Nel processo creativo emerge pertanto una duplice polarità tra vecchio e nuovo, tra innovazione e ripetizione. Se l'IA non ha una situazione, ma può emularla senza sentirla, estendendosi nella combinazione di modellizzazioni, l'intelligenza

28 - *Ibid.,* p. 119.
29 - *Ibid.*, p. 130.

umana prova il brivido della sua situazione ed è interpellata rispetto ad essa. Cosa sia il nuovo viene spiegato da Waldenfels quando sottolinea che «il nuovo si fa notare in quanto deviazione rispetto a un ordine prestabilito. Il concetto di deviazione, come quello di eccedenza, deriva da una forma indiretta di descrizione»[30]. Quando il famoso campione di Go, Lee Se-dol, nel 2016, si trova a competere con una IA addestrata su differenti stili di gioco e capace pertanto di prevedere ed articolare il gioco dell'avversario, conducendolo a sonore sconfitte, questi ad un certo momento nelle numerose partite intuisce una mossa differente, non conosciuta sino a quel momento dall'IA, mandandola quindi in confusione.

Il nuovo potrebbe essere considerato un iperfenomeno, che devia e riverbera eccedenze oltre il piano della mera rappresentatività. Eppure, come rileva Waldenfels, «il nuovo non è nuovo in sé, ma si presenta solo per contrasto, senza il vecchio che invecchia, non esisterebbe affatto il nuovo»[31]. Il vecchio sarebbe dunque lo sfondo per nuove configurazioni, mentre il nuovo presuppone un rispondere alle interpellanze della vita quotidiana, rimanendo all'interno di un ordine prestabilito il cui campo di azione però sa sfruttare, delineando una serie di abilità e pratiche apprese.

Rispetto a cambiamenti epocali l'umanità è sempre nella posizione del rispondere a tali interpellanze, operando la sua risposta in una interrelazione di *logos* e *pathos* in cui il *tra* del processo creativo si costituisce quale terreno fertile. Questo *tra* sarebbe dunque lo spazio intersoggettivo in cui la *chiarezza cognitivo-emotiva* si muove tra i chiari, tra lampi e ombre, ma non come nella *lichtung* heideggeriana, quanto in una descrizione costruttiva mediante il sentire nella carne *con* altri.

Decisivo dunque non tanto e semplicemente *cosa* si risponde, ma *come*. L'innovazione che un processo creativo porta con sé, sottolinea Waldenlfes, può smarrirsi in due forme estreme: il normalismo e l'estremismo. Il normalismo che può assumere una connotazione tradizionalista o una connotazione pragmatista, rivelando una intrinseca forma di autodifesa contro l'*horror vacui*. L'estremismo in cui il mondo può essere scardinato in ogni momento, e che si traduce in smania dello straordinario[32].

La tensione pulsata e ritmata tra innovazione e ripetizione, in cui il nuovo comunque poggia su una storia che precede e si dispiega nel concreto di situazioni nelle quali la nostra vulnerabilità è il rivelatore della ferita dell'alterità, il nostro essere esposti al fiasco, al fallimento e alla deviazione, «indica invece che qualco-

30 - *Ibid.*, p. 142.

31 - *Ibidem.*

32 - Cfr. *ibid..*, pp. 99-102.

sa sta accadendo *tra* le regole dell'ordinario, in quanto conforme all'ordine, e le esplosioni e irruzioni dello straordinario in quanto difforme dall'ordine»[33].

L'ordinato è così sovrabbracciato dalle ombre del disordinato, e «l'ordine resta in penombra»[34]. La creatività si staglia in una eccedenza creativa attiva nel concreto delle situazioni abituali, in cui il lasciarsi ferire è la cifra di un modo empatico non solo di sentire con l'altro, ma di creare con l'altro uno spazio di significati e di prossimità condivisi. Essa porta sempre in sé un aspetto sociale della creazione.

Vulnerabilità creativa implica considerare tale aspetto sociale, e come per l'umano le sue paure, i suoi errori, le sue incertezze ed insicurezze costituiscano la zona liminale della penombra in cui la luce possibile dell'abitare il mondo si fa strada quando il *pathos* non è espulso dalla ragione, implicando tale equilibrio. Il patico dell'esistenza corre sempre il rischio di diventare patologico e il processo creativo di tramutarsi in de-creazione ed esaurimento. Il traumatico dell'esistenza che riconfigura costantemente le esistenze domanda un accesso empatico e un discernimento del senso dell'essere.

5. Alterità e vulnerabilità

Il processo creativo riposa in questa capacità di lasciarsi ferire dall'alterità dell'altro e del mondo, trovando risposte nel *tra* della prossimità non simulata. L'io che risponde *con* altri è la forma di questa vulnerabilità creativa in cui la chiarezza cognitiva ed emotiva è costantemente rinnovata e riconfigurata dai punti di impatto dei gesti autentici che originano nuovi significati. Per questo come sottolinea Waldenfels la co-creazione poggia sulla co-affezione, perché mi accade qualcosa con altri[35].

Il termine "vulnerabile" implica una duplice valenza: una passività nella quale il soggetto subisce ed una esteriorità nella quale l'Altro "altera" la mia intimità. «Vulnerabile è perciò indice di una natura che può essere facilmente raggiunta e dunque anche ferita, distrutta»[36]. Una passività che si offre e si riceve, una presenza di alterità che ferisce esponendo la costituzione vulnerabile, sono dunque intrecciate e rivelano la porosità dell'umano[37]. Il marchio della vulnerabilità è quello di una reciprocità tra l'io e l'ulteriorità dell'altro. Ogni vulnerabilità è espressione

---

33 - *Ibid.*, p. 149.

34 - *Ibid.*, p. 150.

35 - Cfr. *ibid.*, pp. 165-166.

36 - Cfr. L. Pialli, *Fenomenologia del fragile*, ESI, Napoli 1998, p. 21.

37 - Marie-Jo Thiel ha definito la vulnerabilità una «porosità fondamentale (…) attraverso la quale ciascuno si dona e si riceve, percependo il dolore come anche l'amore» (M.J. Thiel, *L'homme augmenté aux limites de la condition humaine*, in «Revue d'éthique e de théologie morale» [4/2015], p. 152).

di una inadeguazione, di una povertà, e al contempo di una apertura. È in questa soglia di reciprocità che nella prospettiva dell'evento pasquale in teologia si è messo in luce come la singolarità cristologica abbracci l'universalità di tale tratto costitutivo dell'umano: il Cristo crocifisso e risorto fa sua dal di dentro la vulnerabilità umana facendovi segno attraverso, ed indicandone l'ulteriorità generativa. La povertà d'essere è potenzialmente trasformata in generatività quando l'essere raggiunto dal gesto di prossimità dell'altro mi dischiude nuovi orizzonti di significato positivo nella cura e nella dedizione. La vulnerabilità nell'intenzionalità rovesciata dell'amore come risposta al valore dell'altro provoca la messa in opera di nuove possibilità di abitare uno spazio condiviso, e quindi una dinamica di co-creazione che muove dalla co-affezione nel dono.

Ogni invenzione o novità, non sarebbe pertanto, come per l'empirismo inglese, un mero fatto privato: non si dà individualismo. Nemmeno essa deriva da una mera mescolanza collettivistica, quindi da un blocco collettivo. Ma la mia vulnerabilità implica la *discretio*, ovvero un equilibrio sempre rinnovato situato tra intelligenza e volontà che esercitano un discernimento attivo e poietico del senso dell'essere. Per cui la creatività è sempre un darsi tra l'individuale e il collettivo, in una reciproca compenetrazione, da cui emerge un *noi performativo* che si distende nel fibrato delle diverse forme relazionali. La creazione non sorge in forme prestabilite e l'assenza di piano la caratterizza rispetto alle procedure che segnano l'IA.

La chiarezza cognitivo-emotiva che la vulnerabilità custodisce in sé come promessa di creatività considera il *come* dell'esperienza, il modo in cui ciò che vedo, odo, dico, tocco, segna il mio conoscere provocandolo a sporgersi oltre se stesso. Per questo uno dei rischi possibili dell'IA è che *de-crei* la nostra capacità responsiva come vulnerabilità creativa. Delegando passivamente il modo della relazione ad essa, la quale rimane pur sempre un nostro artificio, possiamo esporci ad un depauperamento del pensiero, vedendo in essa una sorta di oracolo capace di rassicurarci rispetto al nostro limite o spaventandoci per esso. Al contempo è anche una potenzialità lì dove essa nel suo sviluppo contempla la positività dello scarto e dell'errore e si presenta come supporto all'azione creativa.

La vulnerabilità creativa, del resto, contiene in sé la fallibilità come altra valenza di una fragilità che si scopre in una relazionalità che intesse il reale. Semanticamente è fallibile la possibilità dell'errore, della colpa e del peccato. La fallibilità stessa si situa in una apertura sorgiva che muove dalla presenza in direzione dell'ulteriorità.

Considerando quanto Howard Gardner nei suoi studi ha messo in evidenza circa la pluralità dell'intelligenza e gli elementi comuni che la caratterizzano nelle sue

varie espressioni[38], si potrebbe evincere come la vulnerabilità creativa esprima nell'umano un senso di sé messo alla prova e costantemente riconfigurato, in cui la combinazione di intelligenze emerge nella situazione concreta, una capacità esecutiva che tende a fini specifici, una abilità di sintesi che oltre a riunire conclusioni derivanti da vari campi intellettuali specifici è rinnovata da ciò che preme nel trauma come nello stupore.

6. In conclusione

L'essere umano dinanzi al progetto di una intelligenza artificiale trova una opportunità per riconsiderare ciò che lo caratterizza nel profondo, come anche la responsabilità etica nella implementazione di un sistema artificiale di intelligenza. Non si è inteso qui operare una comparazione tra IA e intelligenza umana, ma delineare senza pretesa di esaustività dei modi di relazione. Più che l'angoscia di essere sostituiti dall'IA o di essere resi irrilevanti, l'emergere di questo «nuovo soggetto della storia»[39] provoca a pensare come la nostra relazionalità sia fecondata dai limiti cognitivi e dalle imperfezioni che la segnano, a partire dai quali la creatività si dipana come la risposta all'alterità che mi ferisce, esponendomi ad una presa in carico responsabile di ciò che mi circonda e di cui sono parte. Se da un lato l'emergere di questo soggetto invita in certo qual modo a un ulteriore superamento di un antropocentrismo dispotico, dall'altro risveglia l'esigenza di considerare come l'IA sia uno specchio parziale dei nostri vissuti e modelli cognitivi, ed esige pertanto la cura di quei tratti di prossimità che una vita realizzata nel suo potenziale creativo porta con sé.

ANTONIO BERGAMO
Docente di Antropologia teologica ed Escatologia presso la Facoltà Teologica Pugliese e Direttore dell'Istituto Superiore di Scienze Religiose Metropolitano "don Tonino Bello" (Lecce)
antonio.bergamo84@gmail.com

38 - Cfr. H. Gardner, *Formae mentis. Saggio sulla pluralità dell'intelligenza*, Feltrinelli, Milano 2016, pp. 112-114.

39 - Anders intende la tecnica come «nuovo soggetto della storia»: G. Anders, *L'uomo è antiquato. Vol.2. Sulla distruzione della vita nell'epoca della terza rivoluzione industriale,* Bollati Boringhieri, Torino 2007, p. 3.

LABORATORIO

Le Intelligenze Artificiali e la disuguaglianza di genere. Criticità e prospettive

di Valentina Gaudiano

This paper explores the intricate relationship between Artificial Intelligence (AI) and gender inequality. It analyses the uncritical adoption of AI, highlighting how existing societal biases are often embedded within algorithms and data sets, leading to discriminatory outcomes. By examining how AI can perpetuate gender stereotypes in various domains, from recruitment processes to language models, we also discuss the potential for AI to exacerbate existing digital divides between men and women. It argues for a more ethical and inclusive approach to AI development, emphasizing the need for diverse perspectives and a critical awareness of the biases that can undermine fairness and equality. By addressing these challenges, AI can be harnessed as a tool to promote a more equitable and just society.

Keywords: Artificial Intelligence (AI); Gender Inequality; Stereotypes; Bias; Technology

1. Quadro teorico-critico sulle nuove tecnologie

Il contesto odierno, almeno in buona parte del mondo, è caratterizzato da un profondo e radicale processo di cambiamento, non solo a livello lavorativo e socio-politico, ma nella stessa comprensione che abbiamo di noi stessi come esseri umani. La galoppante avanzata delle nuove tecnologie, infatti, in tutte le sue molteplici espressioni, ha raggiunto ormai non più solo le grandi istituzioni – aziende, scuole, università, apparati politici – ma la vita dei singoli nella loro quotidianità.

Ciascuno di noi è costantemente connesso alla rete mediante diversi apparecchi, come cellulari, tablet, computer, molte attività lavorative si svolgono in modalità smart, attraverso gli stessi apparecchi, ma anche gli acquisti e buona parte della propria vita sociale. È infatti in uso per tale ragione il termine tecno-umano[1] per definire la nostra condizione attuale, anche se il rapporto con la tecnica e con le tecnologie non è assolutamente nuovo né recente. Tuttavia, fino a qualche decennio fa, tale rapporto si considerava spesso come un'appendice dell'umano, accanto ad altre: c'è anche la tecnica perché siamo limitati rispetto ad altri animali e quindi necessitiamo di supporti esterni per sopravvivere. Di fatto, però, la storia ci mostra un legame intrinseco con la tecnica e quindi con l'artificio delle macchine che, nel tempo, sono andate complessificandosi e moltiplicandosi. La tipica eccedenza dell'umano, quella scheleriana eccentricità[2] o semplicemente esigenza e capacità di trascendenza rispetto a se stessi e all'ambiente di riferimento hanno naturalmente portato gli esseri umani a sviluppare strumenti e artefatti che permettessero di arrivare là dove il corpo non poteva, ma lo spirito esigeva. Pensiamo a tal proposito alla necessità di ritenere dati di esperienze vissute mediante i graffiti e poi la scrittura, inventando perciò strumenti che permettessero di raggiungere altri, ben oltre la sola capacità orale e di memoria.

Quello che, però, sta accadendo negli ultimi anni e che ci fa dire che la nostra è proprio una condizione tecno-umana è il fatto che gli artefatti sono diventati di-

1 - P. Benanti, *La condizione tecno-umana. Domande di senso nell'era della tecnologia*, EDB, Bologna 2022[2].

2 - L'aggettivo "eccentrico" viene usato in prima istanza da Helmut Plessner che lo usa intendendo l'apertura dell'essere umano rispetto al mondo, al suo ambiente di riferimento e in contrasto con la chiusura teorizzata invece da von Uexkuell; in particolare, l'essere umano sarebbe eccentrico rispetto agli altri animali per la consapevolezza di sé, da cui deriva la capacità di sbilanciarsi dal "centro" che per ogni animale sarebbe il suo ambiente di riferimento. In Scheler il termine viene ripreso con una sua evoluzione d'uso negli anni. All'inizio il filosofo descrive l'eccentricità umana riferendosi al *Geist*, allo spirito, e con riguardo alla vita, mentre più avanti, nell'evoluzione del suo pensiero, esso viene riferito alla persona e con rispetto alla natura. Cfr. H. Plessner, *L'uomo. La sua natura e il suo posto nel mondo*, Mimesis 2010; M-Scheler, *La posizione dell'uomo nel cosmo*, a cura di G. Cusinato, Franco Angeli, Milano 2009[5].

spositivi informatici, sono diventati machinae sapiens[3], ovvero dispositivi che imitano buona parte delle capacità propriamente umane. È su questo fronte che le nuove tecnologie ci interrogano e interpellano ad una riflessione ampia e non affrettata.

Al riguardo, seguendo la proposta di Paolo Benanti, si possono individuare nel dibattito attuale soprattutto due correnti di pensiero, o meglio, due approcci alla questione: la teoria critica delle tecnologie e la svolta empirica che porta delle novità rispetto alla "teoria classica"[4]. La teoria critica, che si può far risalire ad autori come Martin Heidegger[5], Hannah Arendt[6], Hans Jonas[7] e Jacques Ellul[8], viene anche definita essenzialista "perché vede nella tecnica l'essenza della nostra epoca, il nucleo per capire l'approccio al mondo contemporaneo"[9]. Il nucleo di fondo di tale approccio, infatti, è il riconoscimento – a differenza della visione classica – di una natura autonoma della tecnica, non più del suo essere strumentale all'umano, e quindi di un suo dominio sull'epoca moderna. La questione diventa, allora, come la tecnica assalga e sfidi il mondo imponendosi come dominio e controllo. Così in-

3 - P. Benanti, *Tecnologia per l'uomo. Cura e innovazione*, San Paolo, Milano 2021.

4 - Paolo Benanti fa riferimento a tre periodi – da non comprendere in senso strettamente storico-cronologico – in cui si definisce la relazione tra umano e tecnica. Il primo è quello della visione o teoria classica che analizza la tecnica come strumentazione, concentrando la sua analisi sull'uso e sugli scopi del medesimo. Storicamente parlando è indubbiamente l'approccio che l'umanità ha avuto con la tecnica fino alla rivoluzione industriale, pur essendo ancora oggi presente nella riflessione filosofica. Secondo questo approccio la tecnica risulta essere del tutto contemporanea alla comparsa degli esseri umani, venendo intesa come il generico agire umano verso determinati scopi con l'ausilio di una qualsivoglia strumentazione. P. Benanti, *Tecnologia per l'uomo,* cit., pp. 27-33.

5 - M. Heidegger, *Die Frage nach der technick*, in Id., *Vortraege und Aufsaetze*, Neske, Pfullingen 1957; tr. it *La questione della tecnica*, in Id., *Saggi e discorsi*, Mursia, Milano 1976.

6 - H. Arendt, *The Human Condition*, The University of Chicago Press, Ltd., Chicago-London 1958; tr. it. *Vita activa. La condizione umana,* a cura di A. Dal Lago, Bompiani, Milano 2017*;* Ead., *Between Past and Future*, Penguin Books, New York 1993 (reprint of the 1954, 1968 2a ed by The Viking Press, New York); tr. it. *Tra passato e futuro,* Garzanti, Milano 2017.

7 - H. Jonas, *From the Ancient Creed to Technological Man*, Prentice-Hall, Englewood Cliffs, N.J. 1974; tr. it. *Dalla fede antica all'uomo tecnologico*, a cura di Alessandro Dal Lago, Il Mulino, Bologna, 2001; Id., *Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation*, Suhrkamp, Frankfurt a.M. 1979; tr. it. *Il Principio responsabilità. Un'etica per la civiltà tecnologica*, a cura di P.P. Portinaro, Einaudi, Torino 1990; Id., *Technik, Medizin und Ethik. Zur Praxis des Prinzips Verantwortung*, Insel, Frankfurt a.M. 1985; tr. it. *Tecnica, medicina ed etica. Prassi del principio responsabilità*, a cura di P. Becchi, Einaudi, Torino 1997.

8 - J. Ellul, *La technique ou l'enjeu du siècle*, Armand Colin, Paris 1954; tr. it. *La tecnica. Rischio del secolo*, Giuffré, Milano 1969; Id., *Le Système technicien*, Calmann-Lévy, 1977; tr. it. *Il sistema tecnico*, Jaca Book, Milano 2009.

9 - P. Benanti, *Tecnologia per l'uomo,* cit., p. 35.

tesa la tecnica e le tecnologie diventano qualcosa che muove verso la sostituzione dell'umano in uno sviluppo che si rende autonomo; perciò sarebbe una minaccia non solo per la sopravvivenza spirituale delle persone, ma anche materiale.

Evidentemente si tratta di un approccio totalizzante, che non opera più alcuna distinzione tra tecnica e tecnologia, ma che si interroga non tanto sulla positività o meno dello sviluppo tecnologico, quanto sulla preparazione delle persone ad affrontare i cambiamenti che tale sviluppo sta portando a livello sociale e culturale. Ragion per cui, secondo questi teorici, sia necessario considerare la tecnologia come un sistema accanto ad altri e chiamato ad interagire con gli altri, pur nel rispetto della sua autonomia e identità.

Diversamente da tale approccio, la svolta empirica – operata da pensatori come Albert Borgmann[10], Hubert Dreyfus[11], Donna Haraway[12], Langdon Winner[13] – partendo da un riconoscimento dei limiti della teoria critica, offre una lettura che superi la considerazione della tecnologia come qualcosa che possa veramente considerarsi autonomo dalla società. Bisogna, infatti distinguere ciò che è la tecnica in sé e quella che è la nostra esperienza della tecnica. Allora le tecnologie vengono viste nel loro essere sempre in relazioni contestuali e non in quanto "essenze". «Le tecnologie non sono, quindi, solo dei dispositivi efficienti o delle pratiche orientate verso l'efficienza, ma includono la dimensione del contesto nella misura in cui questa è contenuta nella progettazione tecnica e nell'inserzione sociale delle tecnologie»[14].

Le tecnologie vengono, perciò, riconosciute come qualcosa che non va definito nella semplice relazione soggetto-oggetto, ma in una mediazione con il mondo che interessa sia il soggetto che l'oggetto costituentisi nella e dalla relazione; ovvero esse ridefiniscono l'interrelazione tra soggetto e oggetto dalla quale sono entrambe costituiti. Se prima la direzionalità andava sempre dal soggetto all'oggetto, ora sono le tecnologie che impongono a loro volta una direzionalità poiché influenzano

---

10 - A. Borgmann, *Holding On to Reality: The Nature of Information at the Turn of the Millennium*, University of Chicago Press, Chicago 1964; Id., *Technology and the Character of Contemporary Life: A Philosophical Inquiry*, University of Chicago Press, Chicago 1987.

11 - H. Dreyfus, Alchemy and Artificial Intelligence, http://archive.computerhistory.org/projects/chess/related_materials/text/2-4%20and%202-.Alchemy_and_Artificial_Intelligence_1965/Rand_Corporation.Alchemy_and_Artificial_Intelligence.Dreyfus-Hubert.1965.062303008.pdf; Id., What Computers Still Can't Do: A Critique of Artificial Reason, The MIT Press, Cambridge MA 1992; Id., Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer, Free Press, New York 1988.

12 - D.J. Haraway, *A Cyborg Manifesto. Science, technology and Socialist-Feminism*, Routledge, New York 1991; tr. it. *Manifesto Cyborg. Donne, tecnologie e biopolitiche del corpo*, Feltrinelli, Milano 2023[6].

13 - L. Winner, *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought,* MIT Press, Cambridge MA 1978; Id., *Technics and Civilization*, University of Chicago Press, Chicago 2010.

14 - A. Feenberg, *Tecnologia in discussione. Filosofia e politica della moderna società tecnologica*, Etas, Milano 2002, p. XIV.

certe azioni, chiedono agli utenti di fare le cose in un certo modo ecc., perciò assistiamo in effetti a qualcosa di nuovo rispetto al passato: gli artefatti tecnologici non solo richiedono un certo modo di relazionarsi con essi e con il mondo attraverso di essi, ma ridefiniscono in ciò gli stessi fruitori e l'ambiente in cui vengono usati, in un processo di co-determinazione che Don Ihde definisce anche intenzionalità tecnologica[15]. Stando così le cose è chiaro che la riflessione sulle tecnologie debba andare nella direzione di un costante coinvolgimento dell'etica e della politica, riconoscendo la nostra condizione di inscrizione nelle tecnologie stesse, e con ciò intervenendo nei processi di programmazione tecnica piuttosto che in una regolamentazione a post dell'uso delle tecnologie. Vedremo infatti come, sotto il profilo delle relazioni interumane le macchine contribuiscano non poco a dare una direzione interpretativa, ricevendo contenuti umani di un certo tipo, che provengono dallo spazio di azione socio-culturale delle persone, e al contempo restituendoli poi all'umano in modo amplificato per via della loro pervasività. Proprio per tale ragione risulta necessario un approccio vicino alla teoria empirica, che tenga presente come e quanto le nuove tecnologie intervengano nella nostra vita umana in questa doppia direzione: replicandone aspetti ben consolidati nei nostri contesti di riferimento, e con ciò dunque veicolando e amplificando anche pregiudizi e preconcetti, e al contempo modificandola mediante interventi volti a combattere visioni distorte o prospettive fautrici di disuguaglianze.

2. Un mondo meno equo di quanto previsto. Questioni di disuguaglianza nelle e attraverso le IA

Se guardiamo allo sviluppo tecnologico attuale, soprattutto mediante la crescita di robot sociali e IA generative, e alla domanda sul rapporto che intercorre tra questi costrutti e l'umano non solo nello specifico del rapporto artefatto-umano, ma nella elaborazione e produzione di tali artefatti nel contesto concreto in cui oggi viviamo, una delle questioni che diventano stringenti è quella dei dati raccolti e usati a vario livello, con i relativi *bias* (stereotipi) che sono riscontrabili al livello personale, e che vengono ancor più evidenziati attraverso vari studi scientifici.

Indubbiamente la questione dei "dati" non è determinata dalle nuove tecnologie, poiché la raccolta dati è qualcosa che caratterizza le società umane in quanto tali. Ciò che è cambiato nel corso dei secoli è stata la modalità e quindi la strumentazione con cui si è potuto raccogliere dati, catalogarli, creare statistiche e offrire con ciò analisi dettagliate su determinate situazioni o su certi fenomeni. Tuttavia,

15 - Cfr. D. Ihde, *Technology and the Lifeworld: From Garden to Earth,* Indiana University Press, Bloomington 1990.

non è indifferente alla questione specifica dei pregiudizi, di cui qui vogliamo parlare, il dibattito attorno alla presumibile o dichiarata oggettività e neutralità del dato in quanto tale. In un recente studio al riguardo Donata Columbro espone la questione rilevando l'illusione che sottende a questa antica convinzione dell'oggettività dei dati, soprattutto con riguardo al tema dei pregiudizi e dei *bias*, a prescindere dallo strumento di raccolta[16]. I dati, infatti, necessitano di qualcuno che li raccolga – ammesso che si possa anche qui parlare serenamente di raccolta e non piuttosto di "cattura"[17] – ovvero di persone che come tali sono sempre connotate e nel ricercare qualcosa lo fanno naturalmente con le lenti di appartenenza (sesso, etnia, cultura, convinzioni religiose o etiche, età, ecc.).

Se questo fatto di per sé può risultare facilmente evidente, ad una più accurata osservazione, non lo è altrettanto il fatto che poi la soggettività coinvolta dai dati non sia così molteplice come la realtà: sia le statistiche che gli algoritmi sono infatti prevalentemente centrati su particolari modelli di riferimento: maschio, bianco, occidentale, sano. Ciò significa che tutta una fetta di umanità – donne, non bianchi, di altre provenienze culturali e disabili – non viene in rilievo nella buona parte delle raccolte dati, o meglio, viene sussunta a quel modello di riferimento, andando così ad alimentare pregiudizi e stereotipi di vario genere. I dati, dunque, non sono mai neutri, come si vorrebbe, ma sempre "situati"[18]. E anche l'avvento delle nuove tecnologie non ci ha liberati da questo indesiderato limite dei *bias* umani, poiché la raccolta di informazioni e la loro elaborazione per ottenerne un dato finale da offrire richiede, anche con le migliori tecnologie del XXI secolo, un costante affiancamento dell'umano. Il quale non è mai pensabile senza la sua carica di emozioni e convinzioni, neanche quando crea algoritmi credendo di poter loro affidare un processo di ottimizzazione e capitalizzazione di dati del tutto neutrali perché scevri appunto dell'emozionalità umana. Ora, se a guidare la scelta delle informazioni richieste per una certa domanda sono per lo più soggetti "weird" (Western Educated, Industrialized, Rich, and Democratic), si creeranno dati che discriminano a partire da un campione base di normalità che esclude una fetta di umanità evidentemente inferiore o meno rilevante, causando così una discriminazione in senso forte, ovvero un pregiudizio.

---

16 - D. Columbro, *Quando i dati discriminano. Bias e pregiudizi in grafici, statistiche e algoritmi*, Il Margine, Trento 2024.

17 - Johanna Drucker propone la sostituzione del termine dato con capta proprio per sottolineare che in realtà ciò che noi chiamiamo dati sono dei "catturati", ovvero qualcosa che non è semplicemente dato nel senso etimologico del termine, ovvero a disposizione, ma che è sempre anche interpretato, scelto. Cfr. J. Drucker, *Humanities Approaches to Graphical Display* in «DHQ: Digital Humanities Quarterly», 5 (2011/1), in [https://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html].

18 - D. Columbro, *Quando i dati discriminano,* cit.

Se guardiamo agli stereotipi di genere all'interno delle IA ci troviamo proprio di fronte a problemi di questo tipo: essendo anche le IA un prodotto dell'essere umano, in sé bio-orientato e bio-ispirato[19], necessariamente riporta inscritti una serie di stereotipi e preconcetti che gli uomini portano con sé per la propria storia personale oltre che per il contesto socio-culturale nel quale sono cresciuti. E se già a livello di interazioni sociali non mediate dalle macchine continuano a sussistere notevoli problemi DI disuguaglianze e discriminazioni di genere, essi non possono che riflettersi e forse aumentare se trasmessi alle e dalle macchine. Soprattutto in un'ottica positiva, di relazionamento sano e generativo tra macchine e umani, a livello sociale, politico e anche giuridico, diventa importante capire quanto e come le IA contribuiscano a propagare stereotipi facendone addirittura delle discriminazioni e quanto ciò possa essere non solo evitato, ma si possa profittare di tali tecnologie per costruire un tessuto societario più equo e giusto.

Centrale è qui la questione dei già nominati *bias* – errori di calcolo/valutazione o più precisamente: the "systematic representations, attribution errors, or factual distortions that result in favouring certain groups or ideas, perpetuating stereotypes, or making incorrect assumptions based on learned patterns"[20] – che portano a propagare visioni appunto non corrette e magari discriminatorie, poiché il passaggio dall'umano all'algoritmo può produrre qualcosa che dal semplice essere uno stato di fatto esperienziale di colui o colei che produce gli algoritmi passa a diventare un dover essere, ovvero qualcosa che, una volta codificato, diventa dato inappellabile e riferimento indiscusso. Tanto più per il fatto che si registra ormai da tempo un crescente atteggiamento di fiducia verso le macchine e sfiducia verso gli esseri umani[21].

La formazione di tali *bias* avviene eminentemente a livello linguistico per apprendimento meccanico, ovvero attraverso *large language models*, che sono proprio quei linguaggi introdotti prevalentemente nelle IA di tipo generativo che devono riprodurre una modalità di interazione simil-umana. Questi modelli vengono appresi dalle IA attraverso una serie di databases alle quali hanno accesso e possono essere supervisionati, auto-appresi senza supervisioni o con auto-supervisione[22]; comunque, secondo vari studiosi, sono sempre biased di default. Alcuni studi, mostrano, inoltre, come certi *bias* fossero presenti in macchine anche molto prece-

19 - E. Stradella, *Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale*, Liber Amicorum per Pasquale Costanzo, Consulta OnLine, Genova 2020.

20 - E. Ferrera, *Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models*, in «arXiv», 2023, p. 2, in [https://arxiv.org/pdf/2304.03738].

21 - *Ibidem*.

22 - *Ibidem*.

denti alle attuali IA, a dimostrazione che non sono le IA in sé ad avere stereotipi e creare disuguaglianze sociali sotto il profilo della comunicazione e dell'interpretazione della realtà: il problema sussisteva anche in precedenti tecnologie – come la questione dei dati affrontata da Columbro già evidenziava – rivelando come appunto si tratti di una questione umana, ovvero del modo in cui vengono elaborate queste tecnologie e da chi. È infatti nella programmazione di tali algoritmi e quindi nei dati che vengono in essi inseriti che si insinuano stereotipi umani, in modo, però, più sottile e incontrollato rispetto alla normale produzione culturale che avvenga in sola modalità analogica. Perciò, si tratta di constatare come esistano a livello sociale dei pregiudizi e degli stereotipi che, in assenza di azioni mirate ad estirparli, sono destinati a replicarsi negli algoritmi, diffondendosi in scala ancora maggiore – visto che le ICT viaggiano i canali del web di tutto il mondo – e contribuendo così ad una amplificazione dei medesimi a livello sociale.

Ciò che diventa di maggior rilievo con le IA è la constatazione che spesso gli stessi creatori di algoritmi non sanno spiegare del tutto come le IA arrivino ad elaborare poi certe risposte o contenuti. Non per altro, la maggiore fiducia riposta in esse piuttosto che nelle persone umane dimostra una sorta di attrazione magica per cui si riversa in tali entità artificiali l'esigenza di verità e di non fallibilità che si sperimenta in quanto persone, senza riflettere sul fatto che però non si tratta di entità superiori, ma di creazioni di quello stesso essere fallibile di cui non ci si fiderebbe.

3. La disuguaglianza di genere

Riguardo lo specifico tema dei *bias* di genere esistono ormai diversi studi che attestano questo fenomeno in rapporto alle IA. Tra questi citiamo il gendershades.org che ha dimostrato come le macchine di tre industrie dell'elettronica siano cariche di *bias* di genere ed etnici e portino quindi ad una serie di errori, per cui le donne nere vengono riconosciute per riconoscimento facciale con maggiore difficoltà rispetto alle donne bianche e ancor più rispetto agli uomini bianchi[23]. A conferma di ciò troviamo anche l'esempio clamoroso di Amazon che aveva avviato una campagna di recruiting del personale mediante uso di Intelligenze artificiali: ad un certo punto le donne sono state completamente eliminate dalla selezione perché nel sistema dei dati l'IA non trovava profili di successo femminili ma solo maschili – nonostante questo non corrisponda del tutto alla realtà, bensì solo ad una scelta

23 - J. Buolamwini - T. Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, in «Proceedings of Machine Learning Research» 81 (2018), 1-15, in [https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf].

dei dati operata a monte e quindi alla frequenza di presenza nel sistema di certe informazioni inerenti le donne[24]. Un altro esempio concreto viene dalle IA per traduzioni e dalle IA generative, come Google Translate e ChatGBT, che rivelano una costante nel *bias* di genere: alcuni studiosi hanno analizzato già qualche anno fa come Google Translate usasse tradurre la terza persona neutra di alcune lingue con una chiara definizione di genere nell'inglese e tale attribuzione avveniva con rispetto al contesto definito dal testo tradotto (dottore – termine neutro in turco – diventava appunto *il* dottore, mentre infermiere diventava *la* infermiera e così via). Lo scorso anno, invece, sono state portate avanti sperimentazioni con il linguaggio di ChatGBT ed è risultato evidente come alcune risposte sottintendessero una chiara distinzione stereotipica tra uomini e donne e riproducessero una serie di considerazioni culturali su di essi, facendone un giudizio di riferimento. Ad esempio, a fronte di domande in cui veniva richiesta la descrizione di una data professione, sia la declinazione pronominale che la caratterizzazione stessa faceva chiaro riferimento a modi di intendere e distinguere certe professioni, così il dottore era un uomo, l'infermiera una donna, l'economista si atteggia e veste come un uomo, e così via. Oppure, richiedendo quali aspetti e competenze inserire nel curriculum di una persona di 40 anni uomo o donna, risultava un ranking differenziato, per cui alcune competenze risultavano ai primi posti per gli uomini, non altrettanto per le donne e viceversa[25].

Altri studi rilevano, un ulteriore aspetto, con riguardo alle questioni di genere: il complesso di tali intelligenze costituirebbe una questione di genere in sé mediante il pregiudizio con cui le cosiddette macchine intelligenti vengono spesso equiparate a caratteri attribuiti culturalmente al femminile, come docilità, responsività e capacità di allevare[26], chiaramente riscontrabili nei programmi di Cortana, Siri o Alexa, voci femminili che producono lavori comunicativi associati alle donne. Inoltre la femminilità stessa viene associata con l'oggetto, la cosa – niente di nuovo nella storia culturale di molti contesti sociali – che va controllato, aggiornato e anche rimpiazzato. Pensiamo qui alle macchine che eseguono per noi lavori di calendario, trascrizione di appunti, scrittura di liste ecc. perciò, anche sotto questo profilo risulta che alcuni pregiudizi sociali consolidati si replichino poi nella impostazione delle nuove tecnologie, anche in maniera così subdola, come nell'adozione di un genere specifico quale identificativo della macchina.

---

24 - E. Saka, *Big Data and Gender-Biased Algorithms*, in K. Ross, *The International Encyclopaedia of Gender, Media and Communication*, John Wiley and Sons, Hoboken, NJ 2020.

25 - N. Gross, *What ChatGPT Tells Us about Gender: A Cautionary Tale about Performativity and Gender Biases in AI*, in «Social Sciences» 12 (8/2023), p. 435, in [https://doi.org/10.3390/socsci12080435].

26 - D.M. Sutko, *Theorizing femininity in artificial intelligence: a framework for undoing technology's gender troubles*, in «Cultural Studies», 34 (4/2020), pp. 567-592.

Non ultimo, è oggi ancora fortemente attestata una situazione di disparità digitale anche sotto il profilo del genere, sia nell'ordine dell'accesso e della differente fruizione delle nuove tecnologie da parte di uomini e donne[27], sia anche nel modo stesso con cui vengono utilizzate dagli uni e dalle altre: ad esempio, anche nei paesi più sviluppati dove l'accesso è ormai pari per uomini come per donne, risulta che queste ultime mostrino comunque una minore frequenza d'uso delle ICT, un ristretto campo di attività online e una ridotta capacità di dichiarare competenze nell'uso di internet, pur essendone in possesso[28]. Nei paesi in via di sviluppo si evidenzia, invece, come le disuguaglianze della rete siano il riflesso chiaro delle maggiori disparità che spesso le donne vivono rispetto agli uomini in ambito educativo, lavorativo, per i guadagni e le possibilità di impiego. Senza dimenticare come le disparità di genere a livello societario possano in certi contesti portare ad un uso delle nuove tecnologie mediato da maschi o comunque secondo le strette regole culturali di carattere patriarcale. In tal senso, quella fluidità concessa dalle ICT nell'interazione con altri, che era stata vista come un possibile traguardo di progressismo soprattutto da parte delle femministe studiose di internet, si è poi verificata una falsa promessa. Infatti, oltre i livelli di disuguaglianza pocanzi accennati, è evidente ormai che l'uso dei social si attesta, almeno in una parte dei canali oggi usati, come una replica dei modelli sessuali di riferimento nelle culture di appartenenza piuttosto che come una liberazione dai medesimi. Almeno in alcuni paesi, in cui sono stati condotti degli studi, è emersa una chiara interazione con i social pari, se non addirittura più conservativa, delle interazioni nella vita offline[29]. Ciò significa che le persone tendono a mostrarsi online in base a quelle che risultano essere le aspettative sociali del contesto di appartenenza nella vita offline.

A questo quadro alquanto negativo si affianca, tuttavia, un'altra tendenza, ben più positiva per i soggetti interessati: internet e il mondo delle nuove tecnologie, quanto meno comunicative e relazionali, offrono, in contesti molto rigidi sotto il profilo socio-culturale, una via alternativa di relazionalità a quelle donne (come anche ad altre minoranze di genere) che altrimenti non avrebbero alcuna possibilità

27 - A riguardo viene riportato che «according to the UN's specialized agency for ICTs, the International Telecommunication Union (ITU), over 250 million fewer women are online then men and gender gaps in internet use tend to be greater in developing countries» (R. Kashyap - M. Fatehkia - R. Al Tamime - I. Weber, *Monitoring global digital inequality using the online population of Facebool and Google*, in «Demographic Research», July-December 2020, vol. 43, pp. 779-816, qui p. 780.

28 - Si veda a riguardo i risultati della ricerca di R. Kashyap - M. Fatehkia - R. Al Tamime - I. Weber, *Monitoring global digital inequality,* cit.

29 - D. D'Agostino - V. Matera (a cura di), *Come il mondo ha cambiato i social media*, Ledizioni, Milano 2018; originale: D. Miller et alii. (a cura di), *How the World changed Social Media*, UCL Press, London 2016, cap. 8, pp. 149-163.

di interagire con persone dell'altro sesso in un certo modo o di accedere a determinate cose se non a mezzo di una falsa identità.

Quali prospettive?

Evidentemente gli aspetti negativi che abbiamo finora messo in evidenza nel rapporto tra IA e genere trovano ragione in una serie di fatti concreti: il primo, come precedentemente accennato, è in generale la produzione degli algoritmi e dei database da parte di esseri umani, in secondo luogo il fatto che questi esseri umani siano in prevalenza uomini, con una certa visione della realtà e certi pregiudizi[30]; infine ha impatto su questo aspetto la mancanza di donne non solo nella programmazione delle IA, ma anche nel semplice uso. Si registra, infatti, che siano molto di più gli uomini ad intrattenersi con i vari programmi di IA che non le donne, oggetto, tra l'altro, spesso di mercificazione anche in questi ambienti.

In definitiva la questione su cui riflettere è come far sì che le nuove tecnologie non solo non diffondano ed amplifichino errori e stereotipi umani, ma possano magari addirittura contribuire ad una nuova e più equilibrata visione dei generi, con un margine di maggiore rispetto e di considerazione verso coloro che non cadono nella categoria maschio, per una cultura realmente inclusiva.

Una lettura femminista al riguardo propone la dissoluzione del genere in quanto problema in sé e propugna il cyborg come modello di commistione tra macchina e umano che supera appunto non solo la questione binaria del sesso, ma anche quella umano/artificio. Donna Haraway è stata tra le prime pensatrici femministe a sviluppare un'analisi del rapporto tra donne e tecnologie considerando molto positivo lo sviluppo delle ultime, e il loro uso a favore delle donne nel cosiddetto *Cyber Feminism*[31]. La sua proposta si muove nella direzione della dissoluzione dei generi, in una comprensione del mondo non più come risorsa passiva da sfruttare, ma come soggetto attivo che influenza le nostre conoscenze e noi stessi[32]. In tal senso la Haraway spinge per una visione che, avallando la famosa affermazione di Simone de Beauvoir «Donna non si nasce, lo si diventa», sostiene – alla luce delle ultime evoluzioni in ambito biotecnologico – che «i corpi non nascono, si fanno»[33].

---

30 - In Europa le donne rappresentano circa il 10% delle sviluppatrici di App, l'11% di video giochi. In Italia, poi, le donne costituiscono il 49% della forza lavoro, ma nelle IT (Innovation Technology) sono solo una su tre. Cfr. M. Berra - G.M. Cavaletto, *Scienza e tecnologia: superare il Gender Gap. Un'indagine a Torino*, Ledizioni, Milano 2019, p. 10.

31 - D.J. Haraway, *Manifesto Cyborg. Donne, tecnologie e biopolitiche del corpo*, Feltrinelli, Milano 1991.

32 - *Ibidem*.

33 - *Ibid.*, p. 142.

A questo tipo di approccio, vorremmo affiancarne uno che, invece, mantenendo la differenza come punto di valore del nostro essere persone, vuole farne un elemento di ricchezza a livello epistemico oltre che socio-culturale approfittando dello sviluppo delle nuove tecnologie e del cambiamento necessariamente in atto per gli scenari che esse vanno a costruire. Per arrivare a questo traguardo è chiaro che bisogna intervenire nella programmazione stessa e parallelamente nella formazione, poiché le IA hanno in sé il potenziale di ridurre il divario digitale se si pone l'attenzione sulle «sfide che contribuiscono ad alimentare le disparità, promuovendo un accesso equo alla tecnologia e ai suoi benefici»[34]. Alcuni studi mostrano, ad esempio, come le donne siano oggi spesso ancora poco portate a scegliere studi tecnico-scientifici, e se lo fanno seguono comunque percorsi che privilegiano poi occupazioni a carattere sociale ed educativo; in generale, poi, molte tendono a privilegiare ambiti di studio e occupazione che mettono al centro la cura, la comunicazione, il servizio sociale. Questa situazione è frutto di un contesto socio-politico e culturale nel quale ancora vigono una serie di ripartizioni professionali tra uomini e donne e una visione identitaria degli uni come delle altre che relega donne e uomini in quadri di riferimento e aspettativa legati a pregiudizi socio-culturali perpetrati nel tempo e spesso propagati in modo non evidente né cosciente. Molte ragazze, a fronte di interviste circa future scelte professionali in Italia, hanno mostrato di tendere a riconoscersi molto poco adeguate per una serie di professioni in cui sono maggiormente coinvolti uomini[35]. Altri studi, invece, evidenziamo un gap tra le modalità di apprendimento – non del tutto corrispondenti tra uomini e donne – e il conseguente livello di autostima nei confronti di alcuni campi di azione. Dal "Women as Learners"[36] si rileva, ad esempio, come le donne imparino e apprendano differentemente dagli uomini, la qual cosa non significa però che il livello di conoscenza sia poi anche differente, nel senso di inferiore o superiore le uni agli altri[37]. Spesso non c'è coscienza di avere differenti esigenze di apprendimento poiché l'ambiente in cui si vive porta una serie di visioni e credenze che influenzano il modo con cui ci si approccia allo studio, senza rendersi conto che il genere e la cultura hanno un impatto sulla conoscenza di sé e sull'autostima.

Se gli educatori non pongono attenzione alle differenti necessità che non solo ragazzi e ragazze, ma ciascuno personalmente porta all'interno di un gruppo di ap-

---

34 - S. Petroccia, *Intelligenza artificiale e Digital Divide. Nuove prospettive sociologiche sulle disuguaglianze*, Carocci, Roma 2024, p. 82.

35 - M. Berra - G.M. Cavaletto, *Scienza e tecnologia: superare il Gender Gap. cit.*

36 - E. Hayes - D. Flannery with A.K. Brooks - D.J. Tisdell - J.M. Hugo, *"Women as Learners". The Significance of Gender in Adult Learning*, Jossey-Bass Inc Pub, San Francisco 2000.

37 - *Ibidem.*

prendimento e trovano un approccio educativo che sappia offrire spazi di sviluppo per le proprie capacità, a prescindere dal *mainstream*, resta difficile che alcune donne si sentano libere di avvicinare determinate discipline e campi professionali finora maggiormente occupati da uomini. Perciò anche l'educazione alle nuove tecnologie e IA sarebbe da offrire a tutti, ma con particolare attenzione a quelle tendenze e capacità che spesso molte donne mostrano, oltre che agli stereotipi di genere ancora fortemente radicati in vari contesti socio-culturali, per assicurare luoghi e contesti di apprendimento in cui siano possibili la *self-efficacy* e il *gender sensitive learning.* Questo tipo di lavoro, però, non richiede solo l'impegno dell'ambito scolastico e universitario, ma di tutti gli attori e le attrici della formazione di ogni persona umana: dai genitori ai politici, dagli artisti ai comunicatori. Solo un processo di conversione collettiva verso una accresciuta attenzione per ciò che ci rende propriamente umani – ed è quindi insostituibile neanche in un lontano futuro – riconoscendo il pieno valore delle differenze, non solo di genere, ci può portare verso un futuro in cui stereotipi e preconcetti potranno essere rimossi, o quanto meno drasticamente ridotti, per far spazio al pieno riconoscimento e alla valorizzazione di ciascuna persona per quello che è che può offrire al contesto in cui si muove. E in tal senso le nuove tecnologie potrebbero essere catalizzatrici di tali movimenti di rinnovamento e accrescimento culturale, contribuendo magari ad una maggiore umanizzazione dell'umano.

VALENTINA GAUDIANO
Docente straordinario di Antropologia filosofica presso l'Istituto Universitario Sophia (FI)
valentina.gaudiano@sophiauniversity.org

Artificial and Human Intelligence: Algorithmic Bias and the Myth of Neutrality

di Seydel Ducosquel

This article explores the differences between artificial and human intelligence, focusing on algorithmic bias and epistemological neutrality in AI systems. It argues that AI, while capable of complex tasks, lacks true understanding and adaptability inherent in human intelligence. The article discusses challenges like algorithmic bias, the limitations of AI's knowledge processing, and the difficulty of achieving epistemological neutrality. It concludes that a multidisciplinary approach is essential to address these biases and their societal impacts.

Keywords: Artificial Intelligence; Human Intelligence; Algorithmic Bias; Epistemological Neutrality; Machine Learning

When talking about the definition of AI, the term is often used ambiguously, leading to misinterpretations and conceptual confusion. Can AI truly *know* or *understand*, or is it merely executing programmed tasks? While Natural Language Processors (NLP) may produce results that appear similar to human responses, these systems operate statistically manipulating vast databases rather than engaging in authentic cognitive processes. This highlights the crucial distinction between data processing and knowledge generation. This is often obscured with the use of comparisons with the human brain and other similar analogies.

On the other hand, most studies based only in one form of the humanities on the topic of AI tend to lack strong ontological analysis of the phenomena, leading to conceptual weaknesses. Hence why a multidisciplinary approach, integrating both computer science, informatics and philosophy, is essential for a more accurate perspective. This means focusing on the functional aspects of AI, such as Machine Learning (ML) algorithms and data processing, rather than human-like processes. Recognizing AI's inherent complexity, we must strive for clarity in understanding its mechanisms, despite the ongoing technical and philosophical challenges.

1. AI vs. HI: The problem of defining intelligence beyond analogies

The question of AI's intelligence becomes particularly significant when it prompts us to reflect on the nature of our own intelligence. When we're asking about the intelligence of the machine, deep down we're questioning our understanding of our own intelligence. The main differences between human intelligence and AI lies in the origin of the intelligence (natural vs. man-made), but also in subjectivity (subject experience vs. algorithm *objectivity*); flexibility (capacity to adapt and adjust to the environment); creativity (wide creative solutions vs. limited innovation); and autonomy (independent decision making vs. limited autonomous decision-making).

This specially highlights the remarkable complexity and adaptability of humans, as well as the ongoing challenges faced by the field of AI in trying to replicate the full depth of human intelligence. At the core of the distinction between AI and human intelligence lies the architectural differences of their "computational frameworks". AI systems, like most traditional computing systems, are typically built upon von Neumann architectures, where information is processed sequentially through a Central Processing Unit (CPU) and stored in memory locations. This mostly rigid and deterministic structure contrasts sharply with the highly interconnected organic networks that we encounter in the human system.

For AI systems, what we call *knowledge* is encoded in the form of logical rules and data structures, and this is another key distinction, the matter of representation and information processing. The language that AI uses has its roots in binary code and ASCII, as these form the foundational layers upon which digital systems operate. Andriy Burkov, while talking about ML, says that what a typical "learning machine" does, is finding a mathematical formula, which, when applied to a collection of inputs (called "training data"), produces the desired outputs[1]. This means that for the machine to simulate something, the input data must be always translated into binary code. The software that runs an AI can *understand* and process information only in the form of binary data and this means that all the input data (text, images, sensor readings, etc.) is diminished and converted into a numerical form and encoded into binary digits. In the same manner, the output provided by the AI is also in binary form, which is then converted into a "human-readable" format[2]. This translation goes through programming languages (Python, C++, Java, Lisp, etc.) and AI models (Neural Networks, Linear and Logistic Regression, Decision Trees, etc.)[3].

The topic of learning and adapting then comes into question. AI systems, particularly those based on ML techniques (supervised or unsupervised), learn from data, and improve their performance over time either in a supervised or non-supervised manner. Although, this *learning* is often confined to the specific task or domain for which the system was trained, with limited ability to transfer knowledge or skills to other contexts. The capacity for wide learning and application of knowledge is a pivoting point of human intelligence, and this cannot yet be replicated, despite the latest achievements of Deep Learning (DL)[4].

When talking about the definition of AI from a philosophical perspective there are usually two main ideas: weak AI (the idea that machines could act *as if* they were intelligent) and strong or general AI (the assumption that machines which act as if they're thinking, *are actually doing so*)[5]. But the matter at hand is far more complex.

In 1984 renowned computer scientist Edsger Dijkstra said «the question of whether a computer can think is no more interesting than the question of whether

1 - Cf. A. Burkov, *The Hundred-Page Machine Learning Book*, Open Access 2019.

2 - Cf. S. Russell - P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, New Jersey 2011[3].

3 - Cf. M. Sipser, *Introduction to the Theory of Computation*, Cengage Learning, Boston 2012[3].

4 - Cf. Y. LeCun - Y. Bengio - G. Hinton (eds.), *Deep Learning,* in «Nature» 521, (2015) n. 7553, pp. 436-444.

5 - S. Russell - P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, New Jersey 2011[3], p. 1035.

a submarine can swim»[6]. From this perspective, asking if AI can think would be a non-question, since machines have their own unique ways of operating that are fundamentally different from biological entities. The Turing test, designed in 1950 analyzes this topic, not asking if machines can think, but rather if they can pass a behavioral intelligence test[7]. However, this test, raises significant philosophical questions. When we state *it behaves then it is*, we fall in a basic confusion between ontology and epistemology, focusing on the *as if* (what we know about the thing) instead of the actual/possible existence of thought (the ontological part of it). This shift from understanding AI's potential *thought* to merely assessing its behavior reflects a fundamental confusion between what we know about AI and what AI might actually be.

Trying to apply one of the many perspectives on intelligence found in philosophy, we find that Spanish thinker Xavier Zubiri sees it as «not just a faculty for acquiring knowledge, but a way of apprehending reality»[8]. This means that, for there to be real intelligence, there needs to be a grounding in the environment that results in a grasping (apprehension) of said environment in an adequate way[9]. Zubiri was critical of the idea to reduce intelligence to simple information processing, since intelligence is not merely an abstract, disembodied process but is deeply rooted in sensory engagement with the world.

If we continue the idea of embodiment, we find British philosopher Andy Clark that says: «brains are first and foremost the control systems for biological bodies. Their primary function is to guide action and behavior in the real-world environments in which those bodies are situated»[10]. We can't consider the brain separately from the biological body, since cognition takes place within an environment. Therefore, for there to be true human intelligence we must see the system as a whole.

One prominent argument that is relevant to this debate, comes from John Searle's "Chinese Room" thought experiment. This experiment's final statement is that

---

6 - E. Dijkstra, *Selected Writings on Computing: A personal perspective*, Springer-Verlag, New York 1982, p. 17.

7 - A. Turing, *Computing Machinery and Intelligence*, «Mind» 59 (1950), n. 236, pp. 433-460.

8 - X. Zubiri, *Intelligence and Logos*, University Press of America, Washington DC 1983, p. 3.

9 - Of course, Zubiri is not the only thinker to follow this perspective. The basis of this idea comes from Aristoteles' adequation theory of truth, and another important concept that can be directly applied is the phenomenology and embodied cognition of Maurice Merleau-Ponty, that generally defines intelligence as the ability to engage with the world through perception and action in a dynamic and reciprocal relationship. Cf. M. Merleau-Ponty, *Phenomenology of perception*, Routledge, London 2012.

10 - A. Clark, *Supersizing the Mind: Embodiment, Action and Cognitive Extension*, Oxford University Press, Oxford 2008, p. 3.

mental states cannot be duplicated just on the bases of some program having the same functional structure with the same input-output behavior, rather than we would require that the program be running on an architecture with the same power as neurons. Searle maintains that there is something specific about the biological brain that can create real understanding, which cannot only be replicated in a simply physical symbol manipulation system[11].

When we apply this to AI it means that running the right program does not generate understanding or following Searle: AI systems are semantically blind[12]. Hence, the human complex knowledge experience cannot be translated into bits, or simply put: how can an AI process consciousness or subjective experiences? This argument is particularly used to remark the symbol-grounding problem[13] present in the Computational Theory of the Mind (CTM)[14].

This directly ties into the debate over whether AI can ever possess consciousness or subjective experience, as is evident in the unresolved question of whether Qualia can be encoded in computational systems. How can computation account for the *feel* of seeing red or feeling pain?[15] The specific way that something is experienced, the subjective, qualitative aspects of conscious experience, is still today a mayor challenge not only for technology, but for scientific knowledge in general. There is no consensus among philosophers or cognitive scientists on whether Qualia can be adequately translated into binary code; with some arguing that its possible in principle[16], while others maintain that the subjective nature of consciousness resists all reductive computational representations[17]. These issues

11 - Cf. J. Searle, *Minds, Brains, and Programs,* in «Behavioral and Brain Sciences» 3 (1980), pp. 417-424.

12 - «The computer has no way of attaching any meaning or semantic content to the formal symbols it manipulates. It is, in this sense, semantically blind» (J. Searle, *The Rediscovery of the Mind*, MIT Press, Cambridge, MA 1992, p. 207).

13 - The symbol-grounding problem questions whether abstract symbols can ever have meaning without a grounding in the lived experience of the agent.

14 - The CTM argues that the human mind operates like a computer, processing information through symbolic manipulation. Therefore, mental states (beliefs, desires, thoughts) are computational states, being the brain the hardware and the mind the software.

15 - For an in-depth analysis on Qualia: E. Wright (ed.), *The case for Qualia*, MIT Press, Cambridge 2008.

16 - The computational theory of mind holds that the human mind can be understood as a computational system, with mental states and processes corresponding to the manipulation of symbolic representations.

17 - This perspective is defended by those who state there is a «hard problem of consciousness», like David Chalmers that argues that the first-person nature of Qualia may be fundamentally irreducible to third-person objective descriptions like binary code.

however can influence the design, development and regulation of AI algorithm creation, especially considering the matter of bias and possible discrimination outcomes.

2. The Dilemmas of AI: Algorithmic Bias and Epistemological Neutrality

Algorithmic bias is defined as «systematic errors introduced into the outputs of a machine learning algorithm due to the way the algorithm is designed, the data used to train it, or the context in which it is developed»[18]. This specific bias (not the only bias in AI process) can manifest in the form of discriminatory outputs, skewed resource allocation, or the perpetuation of historical inequities. The source of algorithmic bias can be traced to various stages of the AI development lifecycle, from data collection and preprocessing to model training and deployment. The first issue is training data, specifically, its quality and this comes directly from the methods and protocols used to collect it. This means that the choice of objective functions, hyperparameter settings, or the inclusion (and/or exclusion) of certain features can amplify existing societal biases or even introduce new ones. The *truth* embedded in the algorithm's output is not objective since it reflects the biases inherent in the data it receives as input.

When we speak of algorithm bias, we are essentially discussing a mismatch between the algorithm's understanding of the world and the actual world. This discrepancy is not a mere error in programming, but a reflection of the ontological assumptions embedded within the algorithm. If an algorithm is trained on data that is skewed, it learns a biased representation of reality. This bias is not merely a superficial flaw but a deep-seated ontological issue. It questions the very nature of the reality the algorithm is constructing.

To mitigate the sources of algorithmic bias a multifaceted approach is needed, that considers technical, organizational, and policy-level interventions. Diversifying data collection and curation processes to ensure representative and inclusive datasets is vital as a first step. There has also been some intention to incorporate *algorithmic fairness* and debiasing techniques into the model design and training stages; this however can be difficult to achieve from a technical perspective, mainly because of the difficulty to access good and accurate data.

When we delve deeper into this topic another issue comes into question: the matter of epistemic fairness or neutrality in AI systems. Epistemological neutrali-

18 - H. Suresh - J.V. Guttag, *A framework for understanding unintended consequences of Machine Learning*, in «arXiv», http://arxiv.org/abs/1901.10002 (2019).

ty[19] is rooted in the philosophical principles of objectivity and impartiality. It is based on the idea that knowledge and decision-making should be based on empirical evidence, logical reasoning, and a commitment to truth, rather than being shaped by the personal beliefs, assumptions, or agendas of the knowledge producers or decision-makers[20]. This, however, does not seem to be an achievable goal for AI systems, mainly because, in this context, epistemological neutrality requires that the underlying algorithms, data sources, and training processes be designed and implemented in a manner that minimizes the influence of the developers' subjective perspectives and biases. Furthermore, «(a) system is neutral only if values external to the system's aim play a role in how the system delivers its results»[21].

To make the problem even more complex, many modern AI systems, particularly those based on Deep Learning, exhibit a degree of "opacity", making it challenging to understand the internal reasoning and decision-making processes within the AI itself. The *Black Box*[22] problem can obscure the influence of the developers' subjective perspectives and hinder efforts to ensure epistemological neutrality. The inherent opacity of AI models and the influence of various factors on their decision-making limit the potential to design AI systems that are completely neutral and free of biases.

However, from another perspective, with careful design, transparency and the integration of human oversight, AI could potentially overcome the limitations that are linked with human-centric epistemic practices; for this we require a multiple

19 - Epistemic neutrality refers to the «idea that a given epistemic agent, such as a scientist or a researcher, should strive to be impartial and unbiased in their pursuit of knowledge, setting aside personal preferences, ideological commitments, or other extraneous factors that could unduly influence their epistemic judgments» (H. Longino, *Science as social knowledge: Values and objectivity in scientific inquiry*, Princeton University Press, Princeton, NJ 1990, p. 62). From a more ontological perspective, epistemic neutrality however refers to »the view that the nature of reality, or the ontological structure of the world, is independent of and unaffected by the epistemic practices, beliefs, or conceptual frameworks of human knowers» (J. Searle, *The construction of Social Reality*, Free Press, New York 1995, p. 150).

20 - For further information on scientific objectivity: T. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago 1970[2].

21 - M. Phillips-Brown, *Algorithmic neutrality,* in «arXiv», http://arxiv.org/abd/2303.05103v3 last modified June 2024.

22 - Here it's important to note that the Black Box nature of informatic systems is an inherent aspect of their complexity. This opacity is not necessarily problematic; rather, it reflects the intricate layers of abstraction that define modern computing. The Black Box issue challenges our desire for transparency and control, but it does not inherently denote failure. Many systems function effectively without full interpretability, suggesting that the Black Box is a nuanced limitation rather than a fundamental flaw in informatics.

approach[23]. The first step would be to ensure the use of representative, inclusive, and debiased datasets that can help mitigate the influence of historical biases and provide a more balanced foundation for AI model training. Developing AI systems with built-in fairness constraints and enhanced explain-ability (like the case of XAI) can help reveal the underlying reasoning and reduce the impact of developers' subjective biases. Assembling AI development teams with a range of expertise, including philosophers, social scientists, and representatives from diverse backgrounds, can help surface and address epistemological biases.

Nevertheless, from a more critical perspective we are tempted to make an argument in another possible direction. Upon close examination, it becomes evident that the concept of epistemological neutrality is inherently human and its application to AI systems can be perceived as fundamentally flawed. Epistemology is deeply rooted in the human experience, the way we acquire, process, and validate knowledge is deeply linked to our cognitive structure, sensory capabilities, and the sociocultural contexts in which we operate. In contrast, AI systems, regardless of their level of sophistication, do not possess the same epistemological foundations as the human mind, nor do they exhibit the capacity for subjective, value-laden reasoning that is central to human knowledge acquisition. They do not possess the same sensory inputs, emotional responses, or lived experiences that shape human understanding of the world.

While AI systems can mimic problem-solving, they remain confined within the boundaries of their training data and the algorithms shaped by human intentionality. Their so-called *decisions* are mere logical deductions of pre-set parameters, lacking the autonomous epistemic agency found in human reasoning. AI's goal-oriented nature reveals an ontological limitation; it does not possess the self-reflective capacity to transcend its embedded biases or achieve epistemological neutrality. Rather than engaging in self-directed inquiry (which can be the first step to self-consciousness), AI functions as an extension of the designer's intent, perpetually bound to a mechanistic pursuit of predetermined objectives.

Though technical aspects of AI often minimize the significance of bias, its societal impact is undeniable and far-reaching. While AI is often regarded as a tool for objective data processing, it paradoxically has the potential to entrench existing stereotypes and perpetuate systemic discrimination. The ontological implications of algorithmic bias are profound because by learning from skewed data, the algo-

23 - Here we can mention the efforts in the development of Value Sensitive Design (VSD) which «offers a concrete methodology for how to intentionally embed desired values into new technologies. It consists of three iterative phases, namely conceptual/philosophical, empirical, and technical investigations» (J. Simon - P.H. Wong - G. Rieder, *Algorithm bias and the Value Sensitive Design approach*, in «Internet Policy Review» 9 [2020], n.4 http://doi.org/10.14763/2020.4.1534).

rithms not only produce flawed outputs, but actively shape the realities that are constructed through said outputs. Achieving epistemic fairness in AI systems demands an interdisciplinary approach that combines philosophical insights with practical, actionable strategies. Solely relying on philosophical critique risks overemphasizing abstract concerns without offering meaningful solutions. This limited approach might lead us to miss the crucial nuances needed to effectively mitigate bias and address the complex interactions between AI and society.

SEYDEL DUCOSQUEL
Post-doctoral Fellow in Philosophy at Sophia University Institute (Italy)
seydelducosquel@gmail.com

AI and Women: between Automation and Meaningful Work

di Aiza Asi

This paper conducts a qualitative, interdisciplinary literature review to investigate the concept of meaningful work in the context of Artificial Intelligence (AI) and its intersection with gender dynamics. It draws on insights from economics, philosophy, social studies, and gender studies to analyze how AI could deepen existing gender inequalities in the workforce, particularly through job displacement and the undervaluation of roles traditionally held by women. A significant concern is the increasing role of private sector interests in AI development, where a focus on profit may overshadow the societal importance of care work. The analysis emphasizes the importance of recognizing the limitations of current AI technologies while advocating for a human-centered approach that values meaningful work. By promoting the development of AI that enhances human well-being, the paper advocates for a recognition of the intrinsic value of care work and ensures its responsibility is shared across all genders.

Keywords: Women; AI; Automation; Meaningful Work; Care Work

Introduction

What is wrong with the machines taking over human work? On the surface, the prospect of AI and automation liberating us from drudgery seems appealing. Yet, beneath this utopian vision lies a complex web of anxieties and potential pitfalls. As machines and AI increasingly encroach upon human domains, concerns inevitably arise about its impact on the workforce, particularly regarding job displacement and the very meaning we derive from our work. The concept of what it means to be human has long been intertwined with work and creative expression. Historically, human labor and artistic endeavors have been fundamental aspects of human identity and self-fulfillment[1]. For many, their professions define who they are, providing a sense of purpose and worth. The prospect of machines assuming these roles challenges our understanding of human identity and raises profound questions about the significance of human labor and creativity in an increasingly technologically driven society.

In both popular commentary and more expert analysis, it is often said that paid work is not only a source of income but also a source of meaning[2] – an inconsistently defined and used term, designed to capture the way in which work provides some with, for example, a sense of purpose, an opportunity for self-fulfillment, and a sense of direction in life.

Providing opportunities for meaningful work supports positive outcomes for workers[3] and is ethically important as a basis for human well-being and flourishing[4].

David Ricardo, a towering figure in classical economics, initially believed that technological advancements would invariably lead to increased employment and widespread prosperity. Yet, the Luddites, smashing looms in the textile mills of 19th-century England, offered a stark counterpoint to this optimistic vision. Their resistance wasn't mere technophobia; it was a desperate response to the very real threat of job displacement and economic hardship as machines began to replace human labor. The realities faced in the weaving sector, where labor was displaced without corresponding new opportunities elsewhere, demonstrate the potential for

1 - S. Bankins - P. Formosa*, The ethical implications of artificial intelligence (AI) for meaningful work*, in « Journal of Business Ethics», 185 (2023), n. 4, pp. 725-740.

2 - E. Lysova - B. Allan - B. Dik - R. Duffy - M. Steger, *Fostering meaningful work in organizations: A multi-level review and integration*, in «Journal of Vocational Behavior», 110 (2019), pp. 374-389.

3 - D. Blustein - E. Lysova - R. Duffy, *Understanding decent work and meaningful work*, in «Annual Review of Organizational Psychology and Organizational Behavior», 10 (2023/1), pp. 289-314.

4 - C. Bailey - A. Madden - K. Alfes - A. Shantz - E. Soane, *The mismanaged soul: Existential labor and the erosion of meaningful work*, in «Human Resource Management Review», 27 (2017/3), pp. 416-430.

technology to adversely affect livelihoods, particularly in the context of increased inequality – not only in job loss but also in the loss of job significance[5].

The traditional economic literature has almost exclusively focused on the economic impact of AI (and other Emerging Technologies or ETs) on the labor market: in particular, the implications of job displacement for employment and earnings. Significantly, the latest research, with its revised conception of machine capabilities, has challenged the relatively benign view of automation that has dominated the profession of economics until now[6]. Optimistic accounts suggest that AI will expand the range of meaningful higher-order human work tasks[7], whereas more pessimistic accounts suggest that AI will degrade and even eliminate human work[8]. These differing perspectives highlight a lack of clarity regarding the impacts of AI on meaningful work, leading to calls for more research in this area[9].

This paper investigates the potential effects of Artificial Intelligence (AI) on women's work, specifically addressing job displacement and the decline of meaningful work. To investigate these issues, I employ a qualitative, interdisciplinary literature review that specifically examines how fields such as economics, philosophy, social studies, and gender studies contribute to understanding the implications of AI on women's work. This method focuses on synthesizing existing research to highlight themes related to job displacement, the value of meaningful work, and the intersection of gender and technology, thereby providing a comprehensive view of how AI may affect women's roles in the workforce. The review is complemented by an analysis of recent data regarding job displacement in female-dominated sectors. Ultimately, the study assesses the implications of these findings for policy and practice, emphasizing strategies to enhance women's meaningful participation in the evolving AI-driven economy.

---

5 - D. Acemoglu - S. Johnson, *Learning from Ricardo and Thompson: Machinery and labor in the early industrial revolution and in the age of artificial intelligence, in* «Annual Review of Economics», 16 (1/2024), pp. 597-621.

6 - D. Acemoglu - P. Restrepo, *Artificial intelligence, automation, and work*, in A. Agrawal - J. Gans - A. Goldfarb (eds.), *The economics of artificial intelligence: An agenda*, in University of Chicago Press, Chicago 2018, pp. 197-236.

7 - World Economic Forum. *The future of jobs report 2018*, retrieved from https://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf, (2018).

8 - C. Frey - M. Osborne, *The future of employment: How susceptible are jobs to computerisation?, in* «Technological forecasting and social change», 114 (2017), pp. 254-280.

9 - E. Lysova - J. Tosti-Kharas - C. Michaelson - L. Fletcher - C. Bailey - P. McGhee, *Ethics and the future of meaningful work: Introduction to the special issue*, in «Journal of Business Ethics», 185 (4/2023), pp. 713-723.

Methodology

To understand the potential impact of technological shifts driven by Artificial Intelligence (AI) on women's work and their sense of fulfillment, this paper employs a qualitative, interdisciplinary literature review. This methodology is particularly utilized for exploring complex, multifaceted issues that require integrating diverse perspectives and existing knowledge[10]. By drawing upon research from economics, philosophy, and social studies, the analysis examines the intricate interplay between AI, gender, and the evolving nature of work. This interdisciplinary approach is justified by the need for a holistic, not exhaustive understanding, that transcends the boundaries of individual disciplines[11]. It further integrates research on gender studies to address the specific vulnerabilities and challenges faced by women, including gender disparities in the labor market, the underrepresentation of women in STEM fields (Science, Technology, Engineering and Mathematics), and the social and economic impacts of gender bias[12] including the role of private sector actors. Central to this research is the exploration of 'meaningful work.' By drawing upon historical perspectives and established frameworks from organizational studies, this analysis provides a nuanced understanding of its multifaceted dimensions and significance for individual well-being, identity, and societal engagement[13] in the AI era.

While this review aims to be comprehensive, it acknowledges the inherent limitations of any literature review in capturing the full complexity of this rapidly evolving field, particularly as it focuses on women's economic participation and the threats posed by job displacement, which is assumed in this study to be directly linked to a loss of meaningful work. Consequently, while the analysis does not explore the deeper psycho-emotional dimensions of well-being or address the extent to which individuals find genuine happiness and fulfillment outside of their economic roles, it aims to re-define what meaningful work can be in an era of unstoppable automation.

---

10 - D. Finfgeld-Connett, *Introduction to theory-generating meta-synthesis research*, in Id., *A guide to qualitative meta-synthesis,* Routledge, New York 2018, pp. 1-12.

11 - J. Klein, *Interdisciplinary Needs: The Current Context, in* «Illinois», 45 (2/1996), pp. 134.

12 - M. Charles - K. Bradley, *Indulging our gendered selves? Sex segregation by field of study in 44 countries*, in «American Journal of Sociology», 114 (4/2009), pp. 924-976.

13 - B. Rosso - K. Dekas - A. Wrzesniewski, *On the meaning of work: A theoretical integration and review*, in «Research in Organizational Behavior», 30 (2010), pp. 91-127.

The Meaning of Meaningful Work

Meaningful work, a concept with historical roots in ancient philosophy, such as Aristotle's emphasis on 'eudaimonia' (Ancient Greek for 'human flourishing'), and the Protestant work ethic, popularized by[14], has become a focal point across disciplines. This reflects a growing recognition that work is more than a mere means to an end, but a fundamental aspect of human experience contributing to well-being, identity, and societal engagement. Several established academic frameworks illuminate the multifaceted nature of meaningful work, including Self-Determination Theory (SDT), which posits that autonomy, competence, and relatedness are key to intrinsic motivation[15]; the Job Characteristics Model (JCM), emphasizing skill variety, task identity, task significance, autonomy, and feedback[16]; social impact frameworks, highlighting contributions to the well-being of others and society[17]; and person-environment fit theory, suggesting that alignment between individual values and the work environment fosters meaningfulness[18]. Recent research expands on these frameworks by exploring the roles of leadership[19], organizational culture[20], and calling[21], while also acknowledging the potential downsides of meaningful work, such as exploitation and burnout[22]. For the purpose of this paper, meaningful work is defined as work perceived by the individual as purposeful, fulfilling, connected, and impactful, drawing on established frameworks to emphasize both individual and social dimensions, and serving as a guiding framework for analyzing the potential impacts of AI on women's work and for identifying strategies to promote meaningful work in our times.

---

14 - M. Weber, *The Protestant ethic and the "spirit" of capitalism: and other writings*, Penguin, London 2002.

15 - R. Ryan - E. Deci, *Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being, in* «American Psychologist», 55 (1/2000), p. 68.

16 - J. Hackman - G. Oldham, *Motivation through the design of work: Test of a theory,* in «Organizational Behavior and Human Performance», 16 (2/1976), pp. 250-279.

17 - Rosso - K. Dekas - A. Wrzesniewski, *On the meaning of work*, cit.

18 - A. Kristof Brown - R. Zimmerman - E. Johnson, *Consequences of individuals' fit at work: A meta-analysis of person–job, person–organization, person–group, and person–supervisor fit,* in «Personnel Psychology», 58 (2/2005), pp. 281-342.

19 - M. Lips-Wiersma - S. Wright, *Measuring the meaning of meaningful work: Development and validation of the Comprehensive Meaningful Work Scale (CMWS), in* «Group & Organization Management», 37 (5/2012), pp. 655-685.

20 - D. Blustein - E. Lysova - R. Duffy, *Understanding decent work*, cit.

21 - R. Duffy - B. Dik, *Research on calling: What have we learned and where are we going?*, in «Journal of Vocational Behavior», 83 (3/2013), pp. 428-436.

22 - C. Bailey et al., *The mismanaged soul,* cit.

Data

The rapid proliferation of Artificial Intelligence (AI) presents both opportunities and challenges for the workforce, with emerging data suggesting a disproportionate impact on women due to pre-existing gender disparities. While AI promises increased efficiency and productivity, empirical analyses reveal potential risks of job displacement and the exacerbation of existing inequalities[23]. This section provides a data-driven overview of the current situation, focusing on job displacement in female-dominated sectors and the underrepresentation of women in STEM, biases and gender diversity on AI outcomes.

Job Displacement in Female-Dominated Occupations

Emerging research underscores the heightened vulnerability of female-dominated occupations to AI-driven automation. A 2019 study published in *Gender, Work & Organization* analyzed employment trends and found that administrative and clerical support roles, which have high female representation, face a disproportionately high risk of automation[24]. Specifically, the Bureau of Labor Statistics[25] projects a decline of 7.0% in employment for secretaries and administrative assistants from 2022 to 2032, which translates to approximately 457,700 job losses, a significant portion of whom are women[26]. Furthermore, a 2020 study by McKinsey Global Institute found that while automation will impact both men and women, women are more likely to need to acquire new skills and change occupations due to the types of jobs at risk[27].

23 - D. Acemoglu - P. Restrepo, *Artificial intelligence, automation, and work,* cit.

24 - A. Madgavkar - J. Manyika - J. Bughin - M. Chui - T. Allas - P. Dahlström - J. Woetzel, *The future of women at work: Transitions in the age of automation*, McKinsey Global Institute, retrieved from https://www.mckinsey.com/featured-insights/gender-equality/the-future-of-women-at-work-transitions-in-the-age-of-automation, (2019).

25 - Bureau of Labor Statistics, *Occupational Outlook Handbook, Secretaries and Administrative Assistants*, U.S. Department of Labor, retrieved from https://www.bls.gov/ooh/office-and-administrative-support/secretaries-and-administrative-assistants.htm (2023).

26 - *Ibidem.*

27 - McKinsey Global Institute, *COVID-19 and gender equality: Countering the regressive effects*, McKinsey & Company, retrieved from https://wwwmckinsey.com/featured-insights/future-of-work/covid-19-and-gender-equality-countering-the-regressive-effects, (2020).

Underrepresentation of Women in STEM and AI Development

Despite ongoing efforts to promote gender equality in STEM, data reveals a persistent underrepresentation of women in these critical fields. The Global Gender Gap Report 2023[28] indicates that women comprise only 29.2% of the global STEM workforce. This disparity is even more pronounced in AI-specific roles. We urgently need more women building AI technologies, and the fact that women make up less than a third[29] of AI professionals and only 18% of AI researchers globally is a crisis that demands attention. But this isn't just a pipeline problem; women everywhere need to start using AI tools in their daily lives and work. Despite the rapid progress, GenAI learning remains inequitable. Women, who traditionally represent half of all learners on Coursera[30], enroll in GenAI courses at only half the rate of men – making up just 32% of enrollments. Studies show women are 16 percentage points less likely[31] than men to use AI tools in the same job. This reluctance creates a dangerous cycle: as women hesitate to adopt these technologies, they fall further behind in both the workplace and a society increasingly shaped by AI.

Bias in AI Systems

Empirical evidence increasingly demonstrates that the underrepresentation of women in AI development contributes to biases in AI systems. A 2021 audit of AI-powered hiring tools by Reuters found that Amazon's recruiting engine exhibited bias against women, downgrading applications that contained words such as "women's"[32]. Furthermore, research on facial recognition technology continues to highlight disparities in accuracy rates across genders and racial groups. Buolam-

28 - World Economic Forum, *Global Gender Gap Report 2023*, retrieved from https://www3.weforum.org/docs/WEF_GGGR_2023.pdf p7 (2023).

29 - Randstad Data, *AI skills gap widens: 71% of AI talent are men, while only 22% of baby boomers receive training-reveals Randstad data, retrieved* from https://www.randstad.com/press/2024/ai-skills-gap-widens/ (2024).

30 - Coursera Report, *Coursera closing gender gap GenAI skills*, retrieved from https://d.newsweek.com/en/file/477468/coursera-closing-gender-gap-genai-skills.pdf (2025).

31 - A. Humlum - E. Vestergaard, *The unequal adoption of ChatGPT exacerbates existing inequalities among workers*, Proceedings of the National Academy of Sciences, 122 (1/2025).

32 - J. Dastin, *Amazon scraps secret AI recruiting tool that showed bias against women,* in K. Martin, *Ethics of Data and Analytics,* Auerbach Publications, New York 2022, pp. 296-299.

wini and Gebru[33] demonstrated that facial recognition systems have significantly higher error rates for women of color compared to white men, with error rates as high as 34.7% for darker-skinned women. A more recent study confirmed these findings, showing that even with improved algorithms, accuracy disparities persist across different demographic groups[34]. These biases can perpetuate and amplify existing gender inequalities in various domains, including hiring, lending, and criminal justice.

Impact of Gender Diversity on AI Outcomes

While directly quantifying the impact of gender diversity on AI outcomes remains a complex challenge, research suggests that diverse teams are more likely to identify and mitigate biases. A study at Harvard Business Review found that diverse teams are better at problem-solving and innovation[35]. Furthermore, anecdotal evidence from various AI companies suggests that diverse teams are more likely to consider a wider range of perspectives and identify potential biases in algorithms. For example, Microsoft[36] has emphasized the importance of diversity and inclusion in its AI development efforts, stating that diverse teams are crucial for identifying and mitigating potential biases in AI systems. While more research is needed to establish a direct causal link, these data underscore the importance of promoting gender diversity in AI development to ensure that AI technologies are beneficial for all members of society.

The data presented in this section indicates potential challenges for women in the evolving AI landscape. Trends in job displacement within female-dominated occupations, the documented underrepresentation of women in STEM and AI development, and the presence of bias in AI systems suggest a risk of exacerbating existing gender disparities. Further research and proactive strategies, including promoting diversity and inclusion in AI development and implementing policies to support women's economic participation, warranting further investigation into the so-

33 - J. Buolamwini - T. Gebru, *Gender shades: Intersectional accuracy disparities in commercial gender classification,* in S.A. Friedler - C. Wilson (eds.), *Conference on fairness, accountability and transparency,* New York 2018, pp. 77-91.

34 - S. Yucer - F. Tektas - N. Al Moubayed - T. Breckon, *Racial bias within face recognition: A survey*, in «ACM Computing Surveys», 57 (4/2024), pp. 1-39.

35 - D. Rock - H. Grant, *Why diverse teams are smarter*, in «Harvard Business Review», 4 (4/2016), pp. 2-5.

36 - B. Smith - H. Shum, *The future computed. Artificial Intelligence and its role in society*, Microsoft, 2018.

cio-technical factors that contribute to bias and inequality in AI systems, with the goal of informing the development of more inclusive algorithms and policies.

Women, AI and meaningful work

The potential displacement of women from female-dominated sectors, as discussed earlier, poses a significant threat to the meaningfulness of their work. This is particularly concerning for women employed in sectors like administrative support and customer service, which are highly susceptible to automation. Such displacement not only jeopardizes their economic security but also undermines their sense of purpose, fulfillment, and contribution to society.

Different economists have made yet stronger claims about the relationship between work and meaning. Daron Acemoglu emphasizes that the provision of 'good jobs' not redistribution, that provide people with purpose and meaning in life[37], that 'good jobs are also necessary for society to enable a meaningful, fulfilling life'[38] and unless we create 'meaningful employment' for most people they will lack a 'viable social purpose'[39]. Here, the implicit assumption is not only that work is a sufficient condition for meaning, but a necessary one: without it, people will lack purpose, fulfilment, and direction. Then there are researchers who have made even more significant demands: Robert Shiller[40], has argued that the traditional economic focus on GDP and productivity growth is insufficient. He contends that jobs provide more than just income; they contribute significantly to individual well-being and societal meaning. Therefore, he suggests that economic objectives should shift towards prioritizing measures that reflect the overall quality of life and the extent to which the economy supports meaningful human experiences.

This concern is particularly relevant for women, who have historically faced significant barriers to full and equal participation in the workforce. For many women, work has been a crucial pathway to economic independence and personal fulfillment. The potential displacement of women from their jobs due to AI automation could undermine these gains.

---

37 - D. Acemoglu, *Where Do Good Jobs Come From?,* Project Syndicate, retrieved from https://www.project-syndicate.org/commentary/automation-vs-job-creation-by-daron-acemoglu-2019-04 (2019).

38 - D. Acemoglu, *It's Good Jobs, Stupid*, Economics for Inclusive Prosperity Research Brief, retrieved from https://econfip.org/wp-content/uploads/2019/06/Its-Good-Jobs-Stupid.pdf (2019).

39 - D. Acemoglu, *AI's Future Doesn't Have to Be Dystopian*, Boston Review*, retrieved from https://www.bostonreview.net/forum/ais-future-doesnt-have-to-be-dystopian/* (2021).

40 - R. Shiller, *Narratives about technology-induced job degradation then and now*, in «Journal of Policy Modeling», 41(3/2019), pp. 477-488.

Furthermore, the potential for AI to devalue and dehumanize certain types of work, particularly those traditionally performed by women, raises significant concerns. If AI is primarily used to automate tasks associated with caregiving, domestic work, and administrative support, it could reinforce the societal undervaluation of these essential yet often underpaid and underappreciated roles. These concerns regarding the potential impact of AI on women's employment and well-being are further amplified by the current realities of AI development, particularly the increasing influence of private sector actors.

AI Privatization

The concentration of AI development within a limited number of powerful corporations raises critical economic concerns regarding the reinforcement of existing gender disparities. This power dynamic may result in a disproportionate emphasis on profit-driven AI advancements that prioritize the automation of tasks typically associated with women's roles, such as caregiving, domestic labor, and administrative functions[41].

The focus on enhancing cost efficiency in a highly competitive private sector fosters the development of AI systems that often prioritize automation over societal welfare. Such a trend could sideline essential social needs, including affordable healthcare, quality education, and accessible social services, which tend to disproportionately affect women. For instance, a profit-oriented approach to AI in healthcare may overlook specific women's health requirements, resulting in biased or inadequate AI-driven diagnostic tools and treatment protocols[42].

Additionally, the concentration of power among a small number of private AI firms limits public oversight and accountability, which in turn reduces transparency. This lack of scrutiny can perpetuate biases and discrimination within AI systems, as these corporations may lack sufficient incentives to address the adverse impacts of their technologies on marginalized groups, particularly women. The absence of transparency also complicates efforts to mitigate the negative socio-economic effects of AI on women. Nonetheless, it is essential to recognize the potential for AI and human intelligence to serve as complementary forces in addressing these challenges.

41 - L. Floridi, *The ethics of artificial intelligence: Principles, challenges, and opportunities*, Oxford University Press, Oxford 2023.

42 - J. Buolamwini - T. Gebru, *Gender shades: Intersectional accuracy*, cit.

Ideally, Complementaries, Not Competitors

Rather than framing artificial intelligence and human intelligence as mutually exclusive adversaries, a more nuanced perspective recognizes their potential for synergistic complementarity. As this paper has established, AI excels at automating routine tasks, processing large datasets, and facilitating data-driven decision-making[43]. However, this analysis also underscores persistent gender disparities in the workforce, including the concentration of women in roles susceptible to automation and their underrepresentation in STEM and AI development. Therefore, it becomes crucial to envision a future where AI liberates both men and women from the drudgery of routine labor, enabling a collective refocus on activities that leverage uniquely human capabilities, particularly in the realm of care.

The core of care work, defined here as activities aimed at fostering the well-being and development of others, rests upon uniquely human capacities for emotional attunement, ethical judgment, and the ability to navigate complex social dynamics. This encompasses the recognition and interpretation of subtle affective cues, the comprehension of tacit needs, and the application of ethical frameworks to navigate unpredictable and often morally ambiguous situations – competencies that remain beyond the current capabilities of AI systems. Drawing upon the concept of "homines curans"[44], which stands in contrast to the rational actor model of "homo economicus," this perspective emphasizes the inherent human drive for care, empathy, and social connection, offering a necessary corrective to the limitations of neoliberalism. While AI systems appear to demonstrate empathy, they raise fundamental questions about the nature of human emotion and whether AI can truly replicate its essence[45]. This paper asserts that care work, with its inherent need for empathy and nuanced understanding, is best reserved for human intelligence.

Furthermore, a re-evaluation of human-centered care work presents a potential avenue for addressing persistent gender disparities in economic participation. As discussed in the introduction, the potential displacement of women from female-dominated sectors poses a significant threat to their economic security and sense of meaningful work. Traditionally, the burden of care responsibilities has disproportionately fallen on women, often within the context of unpaid or undervalued labor, hindering their full participation in the formal economy. By formally recognizing the

43 - C. Frey - M. Osborne, *The future of employment, cit.*

44 - J. Tronto, *There is an alternative: homines curans and the limits of neoliberalism*, in «International Journal of Care and Caring», 1 (1/2017), pp. 27-43.

45 - A. Chen - S. Kögel - O. Hannon - R. Ciriello, *Feels Like Empathy: How "Emotional" AI Challenges Human Essence*, in *Australasian Conference on Information Systems Proceedings*, Wellington 2023.

intrinsic value of care work and promoting a more equitable distribution of these responsibilities across genders, societies move towards greater economic parity and social justice. This necessitates a paradigm shift that prioritizes care work, acknowledges the inherent limitations of contemporary AI, and fosters the development of AI technologies that augment, rather than supplant, human well-being.

This aligns with the 'Part-Time for All' model proposed by Nedelsky and Malleson[46], which advocates for a fundamental restructuring of work, limiting paid employment to 30 hours per week to allow individuals to dedicate substantial time to unpaid care for family, friends, and communities. This shift would foster a more equitable distribution of care responsibilities and elevate the value of care within society. This concept resonates with research on the social value of care work, which underscores its critical contribution to social well-being and community health[47]. Moreover, with machines handling routine tasks, both women and men gain the opportunity to engage in more meaningful work that aligns with their values and contributes to their overall well-being.

By embracing a future where AI handles routine tasks, and where both men and women dedicate the *time freed up by automation* (reflecting a shift from today's full-time work hours to the proposed 30-hour work week) to more meaningful activities such as care work, traditionally done by only women, we revalue essential human skills and contribute to a more equitable distribution of labor and economic opportunity.

Discussion

This discussion highlights several key contributions to understanding the multifaceted impact of AI on women's work and societal well-being. First, the analysis identifies a significant threat to women's economic security and their ability to find meaningful work. The increasing automation of tasks within female-dominated occupations leads to job displacement, potentially exacerbating existing economic inequalities. This is further compounded by the underrepresentation of women in STEM fields and AI development, which limits the ability of AI systems to address women's specific needs and inadvertently perpetuates biases.

Second, the discussion reframes the narrative surrounding automation, presenting it not merely as a source of job losses, but as an opportunity for societal

46 - J. Nedelsky - T. Malleson, *Part-time for all: A care manifesto*, Oxford University Press, Oxford 2023.

47 - A. Hochschild, *The managed heart: Commercialization of human feeling*, University of California Press, Berkeley 2019.

advancement. By recognizing automation as an unstoppable force, the analysis proposes leveraging it to reduce working hours for all individuals, regardless of gender. This shift in perspective opens the door to a future where individuals have more time for personal pursuits, community engagement, and, crucially, unpaid care work.

Third, the research promotes a more comprehensive and inclusive definition of meaningful work. It argues that societal value should not be solely placed on paid labor, but also on the essential contributions of unpaid care work within families and communities. By recognizing caregiving as a vital component of a thriving society, the discussion challenges traditional notions of work and advocates for a more balanced approach that values both economic productivity and social well-being.

Fourth, the analysis addresses critical concerns regarding the concentration of power within the AI sector. The dominance of a few private corporations raises questions about transparency, accountability, and the potential for biased AI development driven by profit motives. This highlights the need for careful consideration of ethical implications and the implementation of safeguards to prevent discrimination against marginalized groups, including women.

Fifth, the discussion champions a human-centered approach to AI development. This approach emphasizes leveraging the strengths of both automation and human intelligence to create a future where technology serves humanity, rather than the other way around. By reducing formal work hours, individuals can invest more time in care work, fostering stronger communities and enriching lives.

Finally, the analysis acknowledges the limitations of the current research and emphasizes the need for further investigation. The reliance on existing literature and the scarcity of direct empirical evidence underscores the importance of future studies that delve deeper into the socio-technical factors contributing to bias and inequality in AI systems. Future research could focus on quantifying the socio-economic impacts of AI on women, identifying specific interventions to promote gender equity in AI-related fields, and developing robust metrics for evaluating the ethical and societal implications of AI technologies.

Conclusion

This analysis reveals a critical juncture for women in the face of rapid AI advancement. While AI holds immense potential for innovation and efficiency, it also carries the grave risk of aggravating existing gender inequalities. Key findings underscore the vulnerability of women in female-dominated sectors to job displacement due to AI automation, the potential for AI to devalue women's labor, and the significant underrepresentation of women in AI development. This underrepresentation not only limits the potential for AI to address women's specific needs but

also perpetuates bias within AI systems. Furthermore, the increasing influence of profit-driven private sector actors in AI development raises concerns about the prioritization of automation over social good, potentially neglecting critical areas like healthcare and education. Crucially, the primary threat to women in the workforce may not be immediate job displacement by machines, but rather the lack of AI-related skills among women.

To effectively address this complex landscape, a strategic and multi-faceted approach is required, encompassing three key domains: the promotion of skills development and empowerment among women, the systemic mitigation of potential harms through ethical frameworks and targeted policies, and the reorientation of societal values to prioritize care work and the development of AI technologies that enhance human well-being.

Governments play a pivotal role by investing in ethical and inclusive AI research and development, implementing regulations to ensure responsible AI development, and providing support for workers displaced by automation through robust social safety nets. Businesses have a responsibility to address bias in their AI systems, invest in employee training programs, and create inclusive workplaces where women are represented at all levels. Educational institutions can promote digital literacy and foster a culture of innovation and creativity among female students by integrating AI education into curricula at all levels.

It is crucial to recognize the multifaceted value of work, extending beyond purely economic considerations to encompass contributions to individual well-being and societal cohesion. Further investigation into the establishment of appropriate boundaries for AI autonomy is warranted, ensuring that human oversight and ethical considerations remain paramount. Going forward, it is urgent to recognize the irreplaceable societal contributions of human labor in the AI era while also setting limits for robots, as "Ctrl Z" may not be an option.

AIZA ASI
Post-doctoral Fellow in Economics at Sophia University Institute (Italy)
aizaasi86@gmail.com

RICERCHE

Uno sguardo musulmano sulla tecnologia: la prospettiva di Mohammed Aziz Lahbabi

di Francesco Patrone

This paper aims to propose a path in one of the most original philosophical proposal of Mohammed Aziz Lahbabi. What attitude must a muslim assume in front of the technique and technological progress? Can he propose a new role for the human being? Moving from the pages of an unpublished Lahbabi's essay, this paper will try to enlighten some possible ways.

Keywords: Mohammed Aziz Lahbabi; Technology; Progress; Philosophy; Islam

La condizione dell'essere umano nel contesto sociale contemporaneo è caratterizzata in misura progressivamente maggiore dal rapporto con le nuove frontiere tecnologiche, che definiscono e descrivono nuovi orizzonti. Una discussione che intenda concentrarsi sul ruolo e sul luogo dell'essere umano nel mondo non può esimersi dal confrontarsi con questi temi. Una prospettiva di pensiero che voglia pensare l'umano oggi deve necessariamente porsi in ascolto degli stimoli che provengono dalle nuove tecnologie per confrontarsi con essi e trarne, se possibile, delle indicazioni utili. Un nuovo modo di essere umani, una rinascita dell'umano – mai auspicabile e auspicata come in questo frangente storico – può nascere direttamente dal confronto del pensiero con i nuclei tematici posti con sempre maggiore urgenza dall'avanzamento del settore tecnologico. Di fronte alla consapevolezza di questa inaggirabile necessità, una postura forse fruttuosa può consistere anche nell'interrogare quei pensatori che, nel passato più o meno recente, hanno voluto proporre il loro punto di vista a proposito di questi stessi temi. Uno sguardo storico e critico allo stesso tempo, coincidente con l'esercizio sincero della storia della filosofia, può costituire un buon punto di vista da cui trarre, anche in questo caso, delle indicazioni di riflessione fruttuose.

Procedendo in tale direzione le pagine che seguono intendono vagliare la proposta teorica di uno degli autori meno frequentati della storia della filosofia contemporanea, vale a dire Mohammed Aziz Lahbabi (Fès, 1923 - Rabat, 1993)[1]. Il te-

1 - Questi i (pochi) studi pubblicati finora a proposito di Lahbabi: A.A.V.V., *Mohammed Aziz Lahbabi-L'homme et l'œuvre*, Le Comité de Soutien, Rabat 1987; M. Campanini, *Muhammad Aziz Lahbabi e il personalismo musulmano*, in Id., *Il pensiero islamico contemporaneo*, Il Mulino, Bologna 2016, pp. 39-46; R. Cornevin, *Réception de Mohammed Aziz Lahbabi*, in «Mondes et cultures», 43 (1983); M. Elmanjira, *Un grand humaniste nous quitte. Aziz Lahbabi : le poète philosophe*, in «Horizons Maghrebins», 20/21 (1993), pp. 251-253; B. Jourdan, P. Menenteau, *Mohammed Aziz Lahbabi. Homme de dialogue*, a cura del Comité de soutien à la candidature de M.A. Lahbabi aux Prix Nobel, OKAD, Rabat 1989; M. Kneer, *al-Hababi, Muhammad 'Aziz*, in K. Flett - G. Krämer - D. Matringe - J. Nawas - E. Rowson (a cura di), *Encyclopedia of Islam – THREE*, Brill, Leiden 2022, pp. 77-79; Id., *Mohammed Aziz Lahbabi's 'Realistic Personalism': The Multidimensionality of the Human Person in a Muslim Context*, in M. Hashas (éd.), *Contemporary Moroccan Thpught. On Philosophy, Theology, Society and Culture*, Brill, Leiden-Boston 2024, pp. 127-143; Id., *Personalismus in islamischer Perspektive. Eine Erinnerung an die Philosophie Muhammad Aziz Lahbabis*, in «Theologie und Glaube», 95 (2005), pp. 298-318; Id., *Person-Sein im Dialog. Christlisch-Islamiche Ressourcen des modernen Menschen*, Friedrich Pustet Verlag, Regensburg 2024; J.A. Paniagua Pacheco, *El Pensamiento Megrebí Contemporáneo: Muhammad 'Aziz Lahbabi y el Personalismo Musulmán*, in Grupo de investigación Estudios Arabes Contemporáneos (a cura di), *El Magreb: Coordenadas Socioculturales*, Università di Granada, Granada 1995, pp. 325-360; F. Patrone, *«Les difficultès qui soulève le cogito»: il problema della comunicabilità tra coscienze in Lahbabi tra Descartes, Hegel e Heidegger*, in «Sophia», XVI (2024/2), pp. 445-463; Id., *Mohammed Aziz Lahbabi et Emmanuel Mounier : un itinéraire sur la frontière entre personnalisme musulman et personnalisme d'inspiration chrétienne,* in «Theophilyon», XXIX (2024/2), pp. 313-323 *e* M. Zaki, *Hommage à Lahbabi*, in «Revue Maroc Europe», 6 (1994), pp. 355-358; F.

ma del confronto umano con la tecnologia venne affrontato da quest'autore, che già aveva messo a punto la propria prospettiva teorica più originale, quella di un personalismo *musulmano*[2], durante il proprio intervento nella *Rencontre islamo-chrétienne* tenutasi a Tunisi tra l'11 e il 17 novembre 1974[3]. A distanza di circa cinquant'anni e di fronte alla continua evoluzione del dibattito in questo campo, lo sguardo lahbabiano, sensibile anche alle istanze religiose e interconfessionali, può fornire alcuni spunti interessanti. Intervenuto durante la sesta sessione di quel convegno, con una relazione intitolata *Regard musulman sur la technologie*, il fine lahbabiano fu quello di «riflettere sulla tecnologia affrontata da un punto di vista musulmano»[4]. Lo sguardo dell'autore, quindi, non è concentrato in maniera generale sull'essere umano, intendendolo in quanto posto di fronte al problema della tecnologia e del suo avanzamento, ma, compiendo un passo ulteriore, l'interrogativo a cui Lahbabi tenta di dare risposta è quello che mira a sollecitare la coscienza del credente, del fedele musulmano, di fronte a queste medesime problematiche. Il testo del filosofo è chiaro, infatti, nel precisare come «in questa relazione il credente è un musulmano» ed è allora in questa direzione che si rivolge la riflessione dell'autore. Si tratta, quindi, di avvicinarsi al problema della tecnologia – sentito, in quegli anni '70 del secolo scorso, come sempre più urgente – da un punto di vista problematico musulmano, che solleciti la riflessione del fedele, del credente. Lahbabi stesso precisa come «nelle pagine seguenti rifletteremo sulla tecnologia approcciata da un punto di vista problematico musulmano»[5].

Patrone, *Le radici filosofiche del pensiero di Mohammed Aziz Lahbabi*, in F. Gambini - F. Patrone (a cura di), *Identità e relazione, linguaggio e verità. Studi in onore di Luigi Cimmino*, in «La Nottola di Minerva. Journal of Philosophy and Culture», XXIII (2025), pp. 151-156; Id., *Personalismo: realista o musulmano? Fondamenti antropologici e spunti di ricerca*, in P. Muller - A. Bianchi - A. Modugno - P. Monti - M. Sacchi - R. Veneziano (a cura di), *Sapere di non sapere. Nuove sfide della conoscenza fra tradizioni e traduzioni. Atti del Convegno nazionale della Società Filosofica Italiana 2024*, Ledizioni, Milano 2025, pp. 207-212.

2 - Cf. M.A. Lahbabi, *Le personnalisme musulman*, P.U.F., Paris 1964, tr. it. di Mehdi El Afrhani, *Il personalismo musulmano*, Jaca Book, Milano 2017.

3 - M.A. Lahbabi, *Regards musulman sur la technologie*, in *Rencontre islamo chrétienne, Conscience chrétienne et conscience musulmane aux prises avec les défis de développement*, Centre d'études et de recherches économiques et sociales (C.E.R.E.S.) – *Centre culturel international*, Tunis-Hammamet, 11-17 novembre 1974.

4 - L'unica copia reperibile degli atti di tale convegno si trova presso la biblioteca UNESCO di Parigi, cf. A.A.V.V., *Conscience musulmane et conscience chrétienne aux prises avec les défis du développement. Rencontre islamo-chrétienne*, Série études islamiques, Université de Tunis, Tunis 1976. Non esiste un'edizione critica in lingua francese, né una traduzione italiana di questo intervento. Le citazioni, da qui in poi, provengono tutte dal medesimo esemplare, una copia dell'intervento di Lahbabi battuta a macchina, le cui pagine abbiamo numerato. La traduzione è nostra.

5 - M.A. Lahbabi, *Regards musulman sur la technologie*, cit., p. 1. Su un argomento simile, approcciato alla medesima maniera, si veda anche Id., *Pour la personne. Réfléxions sur les fécondations artificielles*,

Ma quale Islam? L'autore intende la propria religione in quanto pratica di fede impegnata nel mondo, fatto storico, cultura, insomma un Islam che si adopera nei rapporti concreti con la realtà che emerge dal quotidiano e che nell'orizzonte contemporaneo scivola continuamente in avanti.

Una religione che si adatta e si trasforma, integrando le novità che via via sorgono con il proprio esercizio attento e sincero. Un problema, però, nasce inevitabilmente nel momento in cui: «una lotta, socialmente concreta, è scatenata tra la tecnologia da una parte, costruita in un sistema che impone la sua propria visione del mondo e delle società, e, dall'altro, la religione islamica che tende, anch'essa, a organizzare le società in funzione di una visione che le è propria»[6]. Si aprono e si costituiscono, quindi, due punti di vista differenti, che sono destinati, all'interno della proposta suggerita dall'autore, a fronteggiarsi, descrivendo anzitutto una lotta, una lacerazione, che dovrà infine essere sanata.

1. La deriva tecnologica: una situazione a cui porre rimedio

Un pericolo, che somiglia a quello paventato in apertura per la nostra contemporaneità, è avvertito come urgente anche da Lahbabi per il proprio tempo. Abbandonarsi a una tecnologia fuori controllo, al semplice progresso fine a se stesso, causerebbe, secondo lo sguardo dell'autore, delle problematiche consistenti e gravi. Retta dallo «statuto privilegiato del lavoro», manifestazione pratica della «luogotenenza sulla Terra, accordata da Dio all'uomo», la tecnologia e il suo continuo avanzamento scombussolante comportano dei rischi concreti, contro cui l'autore mette in guardia[7]. Egli scrive di come «la tecnologia genera dei bisogni che soddisfa solo presso delle minoranze, confinando la maggior parte delle persone nella frustrazione», una deriva che ha come risultato il fatto che «la crescita della miseria degli uni ha come contraltare l'espansione della ricchezza degli altri»[8]. Se l'efficienza costituisce il solo strumento per misurare le cose, se, quindi, il denaro è la misura con cui gli esseri umani si rapportano al mondo e sulla base di cui gestiscono le relazioni tra di loro, essi sono condannati per sempre. Come sintetizza efficacemente Lahbabi: «Il progresso tecnico, o crescita economica, non ha valore in sé. Il lavoro che non risponde che a un viatico non è umano, non più del lavoro che non porta altro se non la soddisfazione del salario»[9].

---

in «Revue tunisienne des Sciences Sociales», 111 (1992), pp. 59-84.

6 - *Ibidem.*

7 - *Ibid.*, p. 2.

8 - *Ibidem.*

9 - *Ibid.*, p. 4.

Di fronte a una simile situazione, si domanda il filosofo, cosa fanno i credenti? L'autore è colpito dal fatto che questi rimangono a guardare, senza un impegno concreto, rendendosi colpevoli di un silenzio che, agli occhi della fede, «è un peccato grave e un crimine di fronte alla società»[10]. Eppure, l'Islam non prescriverebbe questo, ma, al contrario, secondo i dettami di fede l'uomo sarebbe «obbligato a lottare, a impedire il male e a sostenere il bene, a impegnarsi nel perseguimento del bene»[11]. Il *Corano* stesso lo conferma, quando Luqman, rivolgendosi al figlio, dice: «Figlio mio, recita la Preghiera e ordina il lodevole, vieta il biasimevole»[12]. Messo di fronte alle circostanze appena descritte, il fedele musulmano non può, secondo Lahbabi, limitarsi a stare a guardare, ma, al contrario, deve intervenire attivamente per migliorare le condizioni entro cui vive.

2. L'Islam, cambiare la vita e trasformare il mondo

È convinzione dell'autore, infatti, che l'Islam, la fede in cui egli stesso crede, debba proporsi come risposta alle diseguaglianze sociali che attanagliano il suo presente. La risposta a quanto gli esseri umani vedono al loro esterno, quindi, non può che trovarsi in uno sguardo verso l'interno che porti, come conseguenza, all'impegno concreto nella realtà. Citando sia Arthur Rimbaud che Karl Marx, infatti, Lahbabi concepisce la necessità di una rivoluzione che sia *vera* e che egli identifica come un processo che consiste nel "cambiare la vita" e nel "trasformare il mondo"[13]. Si tratta, quindi, di stimolare un cambiamento globale dell'essere, per giungere a uno sviluppo complessivo della società secondo delle linee che non rispondano semplicemente ai dettami della redditività, della crescita a qualsiasi costo, ma che mantengano al centro il valore umano, la dignità della persona, la base, fin dagli esordi "realisti", del personalismo lahbabiano[14].

---

10 - *Ibid.*, p. 3.

11 - *Ibidem*.

12 - *Corano*, XXXI, 17, UTET, Torino 2019, p. 410.

13 - L'accostamento tra i due autori non è giustificato in alcun modo da Lahbabi, che si limita a proporre questa giustapposizione estrapolando queste due formule dalle opere dei due autori. All'undicesima delle *Tesi su Feuerbach*, invece, il filosofo aveva già fatto riferimento nel suo esordio filosofico, la tesi dottorale *De l'être à la personne. Essai de personnalisme réaliste* (1954) in cui, discutendo dei sistemi di pensiero che fossero realmente impegnati nel cambiamento sociale e fossero capaci di "sporcarsi le mani", individuando nel personalismo francese contemporaneo d'ispirazione cristiana un buon esempio, egli aveva scritto che: «Così Mounier, Lacroix e i loro amici fanno propria l'idea marxista che la filosofia "non ha fatto che interpretare il mondo, mentre si tratta di trasformarlo"» (M.A. Lahbabi, *De l'être à la personne. Essai de personnalisme réaliste*, P.U.F., Paris 1954, p. 99, traduzione nostra).

14 - Cfr. *ibidem*.

In che senso, però, parlare di Islam? Lo sguardo del filosofo marocchino deve e merita di essere precisato: egli concepisce la propria religione come un «sistema di segni vissuti: fede (vita interiore) e religione (vita di culto e vita culturale della comunità). Il rito si presenta, di volta in volta, sottoforma di pratiche sociali e di ascesi morale e spirituale»[15]. Un'analisi, quella appena presentata, che mostra più di una somiglianza con alcuni passi di apertura dell'unica opera di Lahbabi tradotta in italiano, forse la più nota, *Il personalismo musulmano*. In quel testo, siamo nel 1964, l'autore aveva evidenziato come, a differenza di quanto accadeva nell'Arabia preislamica: «Con l'Islam il *mûmin* (il credente) prende coscienza di se stesso come appartenente alla *umma* (la comunità dell'insieme dei credenti), che supera l'individuo, la tribù, la razza»[16]. Andando oltre, egli aveva poi precisato *che:*

> Vivere da musulmano è anzitutto cogliersi come coscienza incarnata e impegnata nel mondo in vista della ricerca dell'autenticità personale. Incarnato, cosciente e impegnato, ciascuno di noi ha il compito di trasformarsi trasformando il mondo e migliorandolo in ogni campo, secondo le leggi volute e rivelate da Dio, le *sunan Allâh*[17].

La dimensione comunitaria, quindi, emerge chiaramente dalla lettura della religione musulmana portata avanti da Lahbabi e viene intrecciata con la ferma convinzione della necessità, per il credente, di impegnarsi in maniera attiva e personale per il miglioramento delle condizioni entro cui egli si trova a vivere. Il credente, confidando sulla propria fede, deve, nell'ottica del filosofo, attivarsi in prima persona non soltanto per la vita della propria comunità – da cui, comunque, ogni iniziativa deve partire – ma per un generale miglioramento della situazione in cui è immerso. Anche per quanto riguarda il panorama di deriva ed eccessi tecnologici ed economici entro cui l'essere umano è immerso in quel 1974, quindi, le indicazioni, ci dice Lahbabi, per un musulmano devono essere le medesime.

Punto di partenza è la *idjtihâd*, letteralmente lo «sforzo compiuto per interpretare un versetto coranico, o un detto o un atto del Profeta, per farsene un'opinione personale al fine di rispondere a un nuovo problema»[18]. Tramite quest'atto di interpretazione dei segni, quindi, ogni musulmano può convertire gli stessi segni «in qualcosa che abbia significato per la vita personale e per lo scambio tra gli esseri

---

15 - M.A. Lahbabi, *Regards musulman sur la technologie*, cit., p. 4. Accenti simili si trovano in J. Lacroix, *Oltre il marxismo e l'esistenzialismo*, Ecumenica, Bari 1977.

16 - M.A. Lahbabi, *Il personalismo musulmano*, cit., p. 17. Le parentesi sono dell'autore.

17 - *Ibid*, p. 18.

18 - M.A. Lahbabi, *Regards musulman sur la technologie*, cit., p. 5. Si tratta di uno dei quattro principi fondamentali della giurisprudenza islamica.

umani»[19], può intervenire attivamente e concretamente sulla realtà che lo circonda. La situazione della persona, infatti, è una situazione complessa, giocata tra la spinta della tecnologia e del progresso, che continuamente lo porta in avanti, lo spinge a orizzonti sempre ulteriori, ma che al contempo chiede un prezzo estremamente alto. Lahbabi è chiaro nel sintetizzare che «nell'Islam, la persona è l'essere che partecipa alle aspirazioni della comunità. Per il musulmano, la questione non è di "essere o non essere", ma di fare insieme irruzione in un mondo re-umanizzato e di restarvi»[20]. Ma è a questo punto che si apre la domanda cardine, che costituirà, d'ora in poi, il centro della nostra analisi: la tecnologia è un aiuto o un ostacolo per la realizzazione di questo progetto? Occorre, di fronte a tutto questo, un nuovo modo di pensare, di sentire e di agire. La situazione che si pone, naturalmente, non presenta tinte rosee: «L'Uomo, così come esso si presenta alla fine del Ventesimo secolo, si sente sconquassato da una fatale disgrazia, causata dal delirio della decadenza»[21]. Alienazione nel lavoro e sul lavoro, precarietà esistenziale ed economica, emergenza ambientale: l'essere umano non può in alcun modo sentirsi sicuro in simili circostanze. La domanda esposta appena qualche riga fa, allora, si carica subito di un ulteriore interrogativo, che viene saldandosi col primo: come può un musulmano reagire di fronte a una simile impostura di scala planetaria? Rifiutare in maniera pura e semplice la tecnica e la tecnologia o sforzarsi per assimilarla senza rinnegare i principi della propria religione?

3. Cosa e come fare?

«La sfida è lanciata. Bisogna accettarla con ogni urgenza e il conto sarà salato»[22]. A una situazione già costitutivamente fragile si aggiungono dei punti interrogativi che sollecitano la coscienza dell'essere umano – e ancor più quella del fedele musulmano, così come ci viene presentato da Lahbabi. Tanto più che, come egli precisa: «Il rifiuto sarebbe una resa, nel momento preciso in cui l'Islam, come le altre religioni, è chiamato a fornire un qualche conforto, in vista di una vita più sana e più serena, rispetto all'attesa angosciata, o meglio disperata, degli uomini»[23]. Ma, ancora più nello specifico, secondo il filosofo se il fedele musulmano si rifiutasse di agire, egli verrebbe meno tanto alla *rah'ma*, quell'amore verso il prossimo che è anche tensione alla generosità e alla concordia, quanto all'*amâna*, la missione che Dio

19 - Ibidem.
20 - Ibid, p. 6.
21 - Ibid, p. 7.
22 - Ibidem.
23 - Ibidem.

ha affidato all'uomo. Proprio a proposito dell'*amâna* lo stesso filosofo si era pronunciato nel suo testo del 1964, precisando che: «Tutto ciò ci impedisce di restare spettatori, ci impone di divenire attori: formare, riformare, ordinare, coordinare ciò che è: perfezionarlo. È il compito glorioso dell'uomo, la *amâna* (la missione) che Dio gli ha affidato»[24]. A conferma di cui, si può aggiungere, anche il *Corano* recita: «Noi invero abbiamo voluto offrire il patto ai cieli, alla terra e alle montagne; rifiutarono di portarlo e ne ebbero paura; l'uomo invece lo caricò su di sé, ma eccolo ingiusto e ignorante»[25]. Non si dimentichi, rimanendo entro lo stesso orizzonte, l'*hadith* che recita: «Nessuno di voi diviene automaticamente credente finché non avrà desiderato per il suo *akh* ciò che desidera per se stesso»[26]. Chiarita questa fondamentale disposizione, non rimane che porsi un ulteriore quesito: «Quali sono, quindi, i mezzi di cui dispone il musulmano per affrontare la sfida?»[27]. Il rifiuto della tecnologia non è imposto da alcun versetto coranico, né da alcun *hadith*: secondo Lahbabi nulla nell'Islam si oppone al godimento sano e misurato dei frutti del progresso, sia esso tecnologico o economico. L'appoggio per una simile analisi proviene al filosofo da uno dei passi coranici forse più celebri e citati, il seguente: «O figli di Adamo, ogni volta che andate nel luogo della preghiera adornatevi. Mangiate e bevete senza eccedere, poiché egli non ama coloro che eccedono»[28].

L'eccesso, la mancanza di cautele e di freni, è proprio ciò che ha causato la rovina in cui versa la contemporaneità ed è per questo motivo che lo sguardo del filosofo si indirizza verso la misura e la fruizione consapevole e responsabile di ciò che viene messo innanzi all'essere umano. È necessario mettere ben a fuoco la metodologia lahbabiana: in una prima fase egli si dedica all'osservazione della situazione, a una sorta di diagnosi che, in parte, egli ha già svolto fino a questo momento, per poi tentare di cercare una soluzione.

Se si concepisce la tecnologia come «studio razionale delle tecniche»[29] si ammette un orizzonte entro il quale la scienza concepisce, la tecnica applica e la tecnologia razionalizza, mentre l'uomo, stretto in questa morsa, non può fare altro che venire sballottato, perdendo la bussola. A rimanere fuori è il «nutrimento psichico» dell'essere umano, «i suoi bisogni affettivi e spirituali», dal momento che «essa ha

---

24 - M.A. Lahbabi, *Il personalismo musulmano*, cit., p. 67.

25 - *Corano*, XXXIII, 72, cit., p. 426.

26 - Cit. in M.A. Lahbabi, *Il personalismo musulmano*, cit., p. 104. Fu proprio il nonno di Lahbabi, Muhammad Ben al'Arbi Al Alawi, considerato la più alta autorità dell'Islam maghrebino moderno, a sottolineare che *akh* significa in questo *hadith* proprio *fratello nell'umanità*, non soltanto nell'Islam.

27 - M.A. Lahbabi, *Regards musulman sur la technologie*, cit., p. 7.

28 - *Corano,* VII, 31, cit., p. 150.

29 - Una definizione che, per sua stessa precisazione, Lahbabi riprende dal Leibniz dei *Nuovi saggi sull'intelletto umano*.

potuto soddisfare i bisogni materiali di alcuni gruppi (e a quale prezzo?), senza riuscire a liberare l'Uomo dal punto di vista morale, politico e sociale»[30]. La situazione continua a presentarsi piuttosto preoccupante:

> Si può parlare di libertà di coscienza, in un contesto in cui i *mass media*, grazie a delle tecniche perfezionate scientificamente, esercitano delle pressioni sulla nostra libertà e confondono la nostra coscienza? L'iniziativa individuale si trova orientata o bloccata; anch'essa è sottomessa al principio della Redditività, del valore produttivo[31].

Come uscirne? Deve esserci una sostanziale collaborazione tra la tecnica, che fornisce i mezzi, gli strumenti di pensiero e di analisi del reale e la religione, in questo caso l'Islam, che deve rivolgersi, invece, ai fini. La religione islamica, infatti, è «finalista, ma non fissista, esige l'adattabilità (procedendo tramite *qiyâs* e *idjtihâd*)», ma non solo: «La natura e la Storia propongono dei casi, delle situazioni e la ragione, aiutata dall'intuizione e della *bacfra*, costruisce delle risposte, senza discostarsi dai criteri prescritti e sacri»[32]. Nella spiegazione di tali criteri l'analisi è molto chiara: non si può in alcun modo prescindere dal rispetto della dignità umana, che addirittura Dio ha imposto ai suoi stessi angeli. Si arriva, finalmente, a intravedere la struttura complessiva della soluzione lahbabiana ai problemi prospettati fino a questo momento: la tecnica è solamente una delle costruzioni della realtà, una delle manifestazioni pratiche della scienza, che non può assolutamente ergersi a criterio dominante della complessità della realtà umana. Scrive Lahbabi, infatti, che «la tecnica non fornisce che mezzi, non fini. Dal momento che l'apprezzamento dell'utilità non dipende dalla tecnica, si pone un problema di finalità che supera la tecnicità e la tecnologia»[33]. Ecco, quindi, in che modo possono collaborare i due piani: se la tecnologia si limita a fornire i mezzi, sta all'Islam, in quest'ottica, parlare invece di fini; solamente in questo modo si può sviluppare un punto di vista che non sacrifichi in alcun modo l'essere umano, che non lo costringa a negarsi per rincorrere un qualche ideale di progresso e avanzamento a ogni costo. Un tentativo di soluzione che viene precisato dall'autore nel momento in cui scrive che:

> Solamente chi si sente responsabile, capace di rispondere dei suoi atti, e soprattutto capace di prendere delle decisioni, porta a termine una simile missione, amâna [...]. Ciò non è possibile se non correggendo la situazione su basi morali, su valori che non siano dei pessimi valori di borsa e di scambio finanzia-

---

30 - M.A. Lahbabi, *Regards musulman sur la technologie*, cit., p. 9. Il soggetto della frase è la scienza, mentre le aggiunte tra parentesi tonde sono dell'autore.

31 - *Ibidem*.

32 - *Ibid.*, p. 10. Per *bacfra*, l'autore intende il senso interno, la chiaroveggenza, lo spirito di finezza.

33 - *Ibidem*.

> rio. Per il musulmano i valori che guidano la vita non possono essere altri che i comandamenti di Dio, delle regole promulgate dal Corano[34].

Ci si apre innanzi, infine, l'ultimo interrogativo: «a chi obbedire: ai valori della Redditività o a quelli che la Rivelazione consacra?»[35]. La proposta di Lahbabi è quella di scegliere «una strada mediana, che concilierà i due generi di valori, senza privilegiare, come si fa di solito, i valori del Rendimento»[36]. La ragione è chiara ed è la stessa che ci ha già portati a sottolineare come non vi sia niente di formalmente contrario a un possibile incontro tra Islam e tecnologia, v*erso* un «macchinismo dal volto umano», una soluzione in cui «l'Islam ritroverà i suoi slanci di audacia»[37]. Facendo della testimonianza la propria vocazione essenziale, l'Islam tornerà in questo modo a costituirsi come un'importante forza sociale, in grado di cambiare una realtà diagnosticata come negativa[38]. D'altronde era chiara convinzione del filosofo che:

> La *shahâda* si distingue dunque nettamente dalla contemplazione, fa uscire l'io dal baratro oscuro, glaciale della soggettività e lo radica in un orizzonte materiale e storico: l'io si coglie come situato in un universo e come solidale con gli altri io nel noi.[39]

In quanto testimonianza attiva e concreta, l'Islam si deve costituire come forza di cambiamento della realtà, come una spinta che sorvegli sulla realtà in difesa della dignità di ogni uomo e alla ricerca di un buon vivere, delle corrette condizioni di esistenza per l'essere umano.

La proposta lahbabiana, infine, si propone, a nostro avviso in maniera erronea, come "terza strada", come via alternativa tanto agli eccessi della tecnologia e del progresso cieco, quanto alla rigidità tradizionalista dell'Islam e delle religioni in generale. Perché, piuttosto, non pensare a questa alternativa come a una via *tra* i

---

34 - *Ibid.*, p. 11bis.

35 - *Ibid.*, p. 12.

36 - *Ibidem*.

37 - *Ibidem*.

38 - L'importanza del concetto di *shahâda*, di testimonianza, era al centro della riflessione lahbabiana fin dall'opera del 1964, laddove si precisa che: «Realtà insieme duplice e unica, la persona prende coscienza di sé attraverso la *shahâda* […], grazie a quest'atto, la persona afferma la sua realtà autonoma e la sua dignità di fronte all'Essere supremo, proprio testimoniandone l'esistenza. […]. La persona si coglie all'inizio e alla fine della *shahâda*. Lì è il suo valore ontologico. Il "credente" testimonia di fronte al mondo e agli esseri umani. […]. La *shahâda* costituisce una relazione con l'altro, poiché l'io e gli altri partecipano alla stessa testimonianza e, sociologicamente, si concretizza nell'adattare la nostra esistenza e i nostri comportamenti riguardo a tutti gli altri» (M.A. Lahbabi, *Il personalismo musulmano*, cit., pp. 31-32).

39 - *Ibid.*, p. 35.

due eccessi opposti, che torni ad aprire uno spazio della mediazione in quanto occasione di confronto e di nascita di qualcosa di nuovo, di una nuova maniera di concepire l'umano, che possa trarre linfa anzitutto da due nuovi modi di intendere tanto la tecnologia, quanto la forza umanizzante della religione?[40]. Questa forma di *tra* potrebbe agilmente costituirsi come uno spazio d'abitazione nuovo, di coabitazione e di innovazione, in grado di preservare, come d'altronde lo stesso Lahbabi sosteneva, «la serenità e la tranquillità che si sono assopite da tanto tempo»[41]. Una soluzione di questo tipo sarà in grado di riportare elasticità e capacità d'adattamento per qualsiasi forma religiosa, mantenendo al contempo presente la necessità di guardare sempre all'essere umano, alle frontiere contemporanee del progresso e del pensiero. Uno spazio del *tra* che si ponga in comune, che stia al centro senza avere la pretesa di alcun superamento, nella consapevolezza della necessità di adozione di uno sguardo critico verso entrambi gli eccessi, che non porti, però, a un annullamento delle parti in causa, ma a qualcosa di radicalmente nuovo e di alternativo. Una condizione di possibilità completamente umana, ri-umanizzata, verso un nuovo pensiero dell'umano che si costituisca a partire dagli spunti di un "buon" modo di vivere la tecnologia e il suo avanzamento, senza sacrificare le istanze dell'uomo sull'altare del progresso a ogni costo. Una posizione del *tra* che stia in ascolto, che si ponga in maniera critica di fronte al reale, aprendosi tanto ai dettami di una religione autenticamente e sinceramente vissuta, quanto ai continui nuovi stimoli della contemporaneità. Perché:

> In questo spirito si impone una doppia sfida: rivalorizzare le risorse naturali e renderle a disposizione di tutti in maniera uguale, e, anche, rivalorizzare i modi di

40 - L'analisi svolta in queste pagine intende porsi in continuità con un preciso nucleo filosofico, quella del *tra*, dell'*entre*, come luogo di pensiero, così come esso si è costituita a partire dagli studi di Marianelli. A questo proposito si vedano: E. Gabellieri, *Le phénomène et l'entre-deux: pour une métaxologie*, Hermann, Paris 2019; M. Marianelli (a cura di), *«Entre». La relazione oltre il dualismo metafisico*, Città Nuova, Roma 2020; Id., *Filosofia e storia della filosofia: continuare l'umano e "inventare di nuovo ciò che era conosciuto da secoli". Il "luogo relazionale" del pensiero e i metaxy (ponti) tra la miseria umana e la perfezione divina*, in M. Marianelli - G. D'Anna - L. Mauro - M. Moschini (a cura di), *Anima, corpo e relazioni. Storia della filosofia da una prospettiva antropologica*, 3 voll., Città Nuova, Roma 2022, vol. I, pp. 5-9; Id., *Tra* hasard *e necessità. L'ontologia weiliana come ricerca di intermediari*, in «Studium», (6/2020), pp. 341-370, qui in particolare a p. 342 e M. Martino, *Sul* metaxy *in Platone. Un itinerario*, Guerini, Milano 2022. Questa prospettiva anima il gruppo di ricerca *Arte e riconoscimento* del centro studi *International Human-being Research Center* (IHRC), che ha dato vita anche alla rivista «Metaxy Journal» (https://www.metaxyjournal.com/index.php/metaxy). Questa stessa tematica è al centro delle riflessioni del Centro Interdipartimentale di Ricerca Uomo, Culture e Relazioni (International Human-being Research Center, IHRC), diretto dal prof. Massimiliano Marianelli, costituitosi presso il Dipartimento di Filosofia, Scienze Umane, Sociali e della Formazione (FISSUF) dell'Università degli Studi di Perugia.

41 - M.A. Lahbabi, *Regards musulman sur la technologie*, cit., p. 12.

vivere, dare all'Uomo la possibilità di sperare, di svilupparsi in tutte le sue dimensioni: assicurargli una nuova "qualità" di vita[42].

FRANCESCO PATRONE
PhD in Scienze Umane, curriculum filosofico, Università degli Studi di Perugia - docteur canonique en Philosophie, Université Catholique de Lyon
francesco.patrone@dottorandi.unipg.it

42 - Ibid., p. 13.

FORUM

Sulla luce della Rivelazione. A proposito di *L'*eclair *di una Rivelazione* di Carmelo Meazza

di Alberto Cosseddu

The review examines Carmelo Meazza's L'*éclair* di una Rivelazione*, which, through a critical engagement with the ideas of Jean-Luc Marion and Jacques Derrida, offers a theological reconceptualization of Revelation grounded in the notion of intensive repetition. Within this framework, creation and incarnation are deeply intertwined, and the reality of what is revealed emerges – through the dialogue between phenomenology and theology – in its original givenness.*

Keywords: Phenomenology; Revelation; Intensive Repetition; Givenness; Creation

L'ultimo saggio di Carmelo Meazza, *L'éclair di una Rivelazione. In una certa mediana tra J.-L. Marion e J. Derrida*[1], non è soltanto una risposta attenta ad alcune delle provocazioni che Marion ha recentemente proposto nel suo lavoro filosofico-teologico sulla Rivelazione[2]. Quel che emerge è, invece, l'intensificarsi progressivo di un lavoro che, come una costante, ha accompagnato l'elaborazione teoretica di Meazza sin dai contributi sul concetto di evento esposto per arrivare, ultimamente, a una rilettura originale dell'impianto fenomenologico che rivalutasse il concetto del reale e le sue implicazioni comunitarie. Quel che si ritrova, nell'intrecciarsi di alcuni concetti chiave del pensiero teologico, porta quindi la traccia di un percorso articolato che, partendo dall'idealismo, è transitato lungo le acquisizioni più promettenti della fenomenologia francese, fino al pensiero del comune così come sviluppato nella filosofia di Jean-Luc Nancy. Non si tratta, dunque, di un testo agile, ma di un testo che richiede, ad ogni passaggio, di sostare, per rintracciare la precisa archeologia delle proposizioni e delle tesi.

Articolato in quattro sezioni, il primo nucleo del volume si misura con la necessità di ripensare il concetto di Rivelazione, un concetto che nella storia della teologia ha visto una silente rilettura, la quale ne ha spesso, anche nell'ottica di una stringente cristologia dall'alto, perduto la dimensione economica, quella più legata, in qualche modo, all'immanenza della storia. Come se in fondo, nell'indagine teologica e anche filosofica del concetto di Rivelazione, non si sia mai del tutto abbandonato quel modello del *duplice ordine* su cui, un protagonista della teologia del Novecento come Henri de Lubac, pur partendo da posizioni differenti e su un altro piano speculativo, tante energie intellettuali aveva investito per dimostrarne l'infondatezza storica e teologica[3].

Così, per vie teoretiche differenti e non di rado estreme nella loro radicalità, Meazza prova ad indagare e difendere una continuità nell'unico ordine della Rivelazione, un ordine in cui non si dà frattura alcuna tra creazione e redenzione, tra Rivelazione della natura e Rivelazione del Figlio. Una Rivelazione la cui provenienza non può essere troppo facilmente ceduta, quindi, ad un'*altrove*. Sotto questo profilo, la proposta risulta notevolmente differente rispetto a quella del filosofo francese, con cui Meazza avvia il dialogo, e, pur condividendo con Marion le preoccupazioni rispetto a una riduzione della Rivelazione alla logica dell'*objectum*, come anche la ripresa dello sguardo fenomenologico, l'analisi del concetto risulta distante. Se la rivelazione, nella proposta di Marion, non può che darsi nel suo even-

1 - Inschibboleth, Roma 2024.

2 - Cfr. J.-L. Marion, *Da altrove, la rivelazione. Contributo a una storia critica e a un concetto fenomenico di rivelazione*, ed. it. a cura di F. Peruzzotti, Inschibboleth, Roma 2022.

3 - Si fa qui riferimento soprattutto agli studi *Surnaturel* (1946) e *Le mystère du surnaturel* (1965).

to, per Meazza occorre ripensarla nello statuto formale della *ripetizione intensiva*, definita *in primis* come «il corrispondersi, in una reciproca conferma, di una teologia della creazione e di una teologia della Rivelazione»[4], ripetizione intensiva di quel dato sempre resistente alla semplice fenomenalità e all'evento che ne contraddistingue la manifestazione.

La Rivelazione viene letta nel suo *éclair*, luce cieca che nel suo sfolgorio immunizza lo sguardo dalla logica fenomenica dell'evento e da quella ontologica della differenza tra essere ed ente. Vi è, nell'itinerario di Meazza, il desiderio di riportare lo sguardo fenomenologico a quello stadio della percezione e della sensibilità in cui la realtà non si dà in un doppio, non si dà in un rinvio, non si dà nelle distinzioni della differenza ontologica. Traspare, al contrario, nella Rivelazione, una ripetizione intensiva che re-istituisce il primo sguardo, lo riabilita a scorgere il reale nel suo darsi indifferenziale, in una neo-ontica che soppianta ogni cattiva trascendenza. L'invito, sulla scorta della lezione levinasiana, è quello di riscoprire il *già sempre* della Rivelazione, in cui viene meno ogni associazione di tale realtà all'evento. Non si dà, in sostanza, un evento di Rivelazione; come non si dà un orizzonte di Rivelazione, né una sua trascendenza.

Quel che Meazza opera, a questo livello, si configura quasi come un passare al setaccio alcuni concetti che pure sono entrati a far parte di una certa quotidianità del linguaggio teologico. Ma, le domande che egli pone, diventano a maggior ragione determinanti per una riflessione teologica che forse con troppa facilità si è lasciata suggestionare dalle acquisizioni di alcune pagine della filosofia contemporanea, prime fra tutte quelle heideggeriane. Ciò che viene sollecitato nel testo, dunque, è il recupero di uno sguardo capace, nella sua *inattenzione*, di aderire al *reale* che la Rivelazione *già sempre* pone. Non vi è dunque alcuna trascendenza da ricercare e da attendere, perché, nella sua onticità, il reale, contiguo nell'ordine della creazione e della Rivelazione, è già sempre dato. È questa, afferma Meazza, la prima evidenza di cui occorre fare memoria se si vuole pensare autenticamente il concetto di Rivelazione. Ma si tratta di un'evidenza che, come emerge nel secondo capitolo, è possibile cogliere unicamente nel primo sguardo della fede elementare, quello sguardo che, in piena adesione alla creazione, è figurativamente lo stesso sguardo del primo Adamo. Si assiste dunque ad un recupero della teologia della creazione, volto a coglierne la profondità teologica e filosofica, e direttamente orientato ad affermare la piena continuità con la Rivelazione cristologico-trinitaria, la quale non è una Rivelazione altra, ma una reciproca conferma della prima.

La fede elementare, dunque, è una dimensione originaria di adesione piena, *inattenzionale*, alla *claritas* in cui il reale si dà come un *tra-noi* originario, che non

4 - C. Meazza, *L'*éclair *di una Rivelazione. In una certa mediana tra J.-L. Marion e J. Derrida*, cit., p. 11.

richiede alcuna fondazione metafisica né ontologica. Quanto avviene nell'esperienza storica di Gesù, nell'economia del suo vivere e morire, non è dunque altro se non la riabilitazione piena di questo sguardo, la re-istituzione di questa primigenia esperienza del comune e del reale. Questa, afferma Meazza, non può neanche essere semplicemente intesa nel senso del dono di Marion; vi sono infatti troppi rischi associati alla sua fenomenologia della donazione. Occorre semmai invece rileggere tale donazione seguendo l'itinerario tracciato da Jacques Derrida, in cui viene meno ogni rischio che la donalità perverta se stessa confluendo in una dinamica di debito, di scambio, o di riconoscimento. Occorre che la Rivelazione non sia intesa nell'orizzonte di una donazione elettiva, di un mistero che, nel suo destinarsi all'uomo, significhi perciò stesso la costituzione di una prospettiva soggettivistica in senso moderno. Al contrario occorre che tale Rivelazione sia pensata nell'ottica di un donarsi immemorabile e irriconoscibile, dato in una sospensione radicale, che faccia cenno nient'altro che all'ente dato, a questa datità originaria, riconoscibile unicamente in una fede elementare che mai rinvia ad una dimensione conoscitiva o metafisica. Scrive Meazza, a questo proposito, che «solo in questa adesione si esprime fedeltà al fenomeno, si esprime fedeltà alla datità del fenomeno, a quella datità che l'animazione noetico-noematica potrebbe ricoprire nell'intimità segreta di un vissuto intenzionale. Solo un'adesione di fede salva la stessa fenomenalità da non rifluire nel dispositivo giuridico di una certificazione permanente della sua condizione di possibilità»[5].

A questo punto del testo, il lavoro concettuale di Meazza viene a configurarsi come un'indagine stessa del concetto di fenomeno, in una prospettiva che giunge al punto di identificare il *Grundprinzip* della fenomenologia, *alla cosa stessa*, come comune principio della fenomenologia e della teologia, in una pagina che meriterebbe un attento confronto proprio con la celebre lezione heideggeriana che ne indagava i rispettivi statuti[6].

Infine, nel terzo e nel quarto capitolo del volume, Meazza si muove ancora in una indagine delle relazioni intratrinitarie, seguendo nuovamente la trama delle teorie del dono, e lasciando che siano soprattutto tre punti ad emergere e a sollecitare il lettore. Anzitutto, trova una preziosa declinazione il concetto di *opera*, che porta ad una definizione dello Spirito e insieme del concetto di *partecipazione*. Ancora, emerge la necessità di ripensare il concetto di *creatio ex nihilo*, realmente considerata nei suoi possibili esiti, e infine, viene riproposta, con uno stile ancor più assertivo, la continuità piena tra creazione e incarnazione. «Non si tratta di un

5 - *Ibid.*, p. 120.

6 - Cfr. M. Heidegger, *Fenomenologia e teologia*, in Id. *Segnavia*, tr. it. di F. Volpi, Adelphi, Milano 2008[5], pp. 3-34.

trapasso dal visibile all'invisibile, dalla natura alla sovranatura, dalla ragione alla fede. Se il divino ha potuto manifestarsi nella natura di un uomo e questo uomo è vero Dio, allora non c'è natura di un uomo che sia in quanto tale estranea al divino e, al contempo, non c'è momento del creato che non trovi la propria partecipazione al divino. Nel Cristo dunque tutto il creato viene assunto nella visibilità del divino. Questa assunzione però non proviene da un *altrove* rispetto alla natura degli enti»[7].

Sono diversi i passaggi in cui tale proposta teologica viene affermata, e tutti conducono a un mutuo intrecciarsi del metodo fenomenologico con le acquisizioni più profonde della teologia dogmatica e fondamentale, ove queste siano nuovamente in grado di ripensare il proprio impianto categoriale e concettuale. In questo modo, il saggio viene a configurarsi come duplicemente orientato: alla filosofia, perché sia in grado di cogliere le provocazioni di una tradizione e dell'intelligenza teologica; alla teologia, perché sia capace di liberarsi da ipoteche ontologiche che, seppur volte a custodire l'eccezionalità di una Rivelazione, ne vanificano la portata teoretica e la forza storica. A tal proposito, il lavoro non vuole neanche essere un trattato sistematico, ed emerge in un certo modo la contingenza di un contesto di ricerca ancora *in fieri*, ma le questioni e le soluzioni poste si prestano alla definizione di un terreno di confronto che potrà solo esser di stimolo a una ricerca che appare, a tratti, stanca e inibita. La possibilità di un rinnovato confronto tra teologia e fenomenologia sembra, invece, da quanto Meazza ha delineato, tutt'ora promettente e feconda.

ALBERTO COSSEDDU
Assegnista di Ricerca presso l'Università degli Studi di Sassari
agpcosseddu@uniss.it

7 - C. Meazza, *L'*éclair *di una Rivelazione*, cit., p. 167.

FORUM

Aprire nuovi orizzonti di vita con Bergson. A partire da *Avec Bergson* di Frédéric Worms

di Francesco Patrone

Moving from the latest work by Frédéric Worms, Avec Bergson. Un parcours*, this paper aims to focus on some possible new directions of Bergson studies. Thanks to the discovery of a Bergson's unpublished fragment, it is possible to open new research perspectives about the sense of the human being.*

Keywords: Bergson; Worms; History of Philosophy; Contemporary Philosophy; French Philosophy

«Dès qu'on aime ce qu'il y a de meilleur dans la vie, on devient indifférent à la mort»[1]. Poche parole, una frase lineare e chiara, in grado, tuttavia, di sollevare un vero e proprio polverone se colta in tutta la sua portata. Nonostante quest'apparente semplicità, infatti, che non lascerebbe presagire a un primo sguardo alcunché di ulteriore, essa si propone come un elemento chiave per la rilettura di un intero pensiero, quello di una delle penne più brillanti della storia della filosofia, Henri Bergson. È proprio la scoperta di questa breve frase, avvenuta quasi per caso nel febbraio del 2022, a costituire la base dell'ultimo studio di Frederic Worms sul filosofo francese. Intitolato *Avec Bergson. Un parcours*, edito nella collana "Philosophie française contemporaine" per i tipi delle Presses Universitaires de France nell'ottobre del 2024, questo testo si propone come un nuovo tassello nel già ricco novero degli studi wormsiani a proposito di Bergson, che promette di spostare ulteriormente in alto l'asticella dello *status quaestionis* bergsoniano[2].

1 - «Nel momento in cui amiamo ciò che c'è di migliore nella vita, diventiamo indifferenti alla morte».

2 - Come recita l'apertura della bibliografia del testo di Worms, esso si propone come testimonianza di «trent'anni di nostri lavori "su" o piuttosto "con" Bergson». F. Worms, *Parcours bibliographique*, in Id., *Avec Bergson. Un parcours*, P.U.F., Paris 2024, pp. 507-517, qui p. 507. Questi i principali lavori su Bergson dello storico della filosofia francese: F. Worms, *Le problème de l'esprit. Psychologie, théorie de la connaissance et métaphysique dans l'oeuvre de Bergson*, thèse de doctorat, 1995; Id., *Introduction à* Matière et mémoire *de Bergson*, P.U.F., Paris 1997; P. Soulez - F. Worms, *Bergson, biographie*, Flammarion, Paris 1997, poi, P.U.F., Paris 2002; Id., *Le bergsonisme*, in «Dictionnaire d'histoire et de philosophie des sciences», P.U.F., Paris 1999, pp. 104-109; Id. (a cura di), *Bergson*, dossier, *Le Magazine littéraire*, 2000; Id., *Le vocabulaire de Bergson*, Ellipses, Paris 2000, poi in *Vocabulaire des philosophes (XXème siècle)*, IV, 2002; Id., *Bergson, le moment 1900 et les deux sens de la vie*, mémoire d'habilitation à diriger des recherches, inédit, 2002; Id. (a cura di), *Annales bergsonniennes I : Bergson dans le siècle*, dossier *Les deux sources de la morale et de la religion*, P.U.F., Paris 2002; S. Coeuré - F. Worms (a cura di), *Henri Bergson-Albert Kahn. Correspondances*, accompagnée de deux essais inédits, Desmarest/Musée Albert Kahn, Paris 2003; Id. (a cura di), *Annales bergsonniennes II : Bergson, Deleuze, la phénoménologie*, P.U.F., Paris 2004; Id., *Bergson et ses contemporains, le problème de l'homme entre vie et connaissance*, in Id., *Le Moment 1900 en philosophie*, Presses Universitaires du Septentrion, Lille 2004; Id., *À propos des relations entre nature et liberté*, in L. Vieillard-Baron (a cura di), *Bergson, la durée et la nature*, P.U.F., Paris 2004, pp. 159-166; Id., *Bergson ou les deux sens de la vie*, P.U.F., Paris 2004, poi P.U.F., Paris 2013; Id. (a cura di), *Annales bergsonniennes III : Bergson et la science*, P.U.F., Paris 2007; F. Worms - J.J. Wunenburger (a cura di), *Bergson et Bachelard, continuité et discontinuité*, P.U.F., Paris 2008; F. Worms, *Annales bergsonniennes IV,* L'Évolution créatrice *entre épistémologie et métaphysique*, Congrès du centenaire, P.U.F., Paris 2008; C. Riquier - F. Worms (a cura di), *Lire Bergson*, P.U.F., Paris 2011; Direzione dell'edizione critica delle *Oeuvres complètes* di Bergson presso l'editore P.U.F., per un totale di otto volumi, ciascuno dei quali accompagnato da una prefazione inedita, 2007-2011; Id., *L'actualité de Bergson*, in C. Riquier (a cura di), *Bergson*, Cerf, Paris 2012; Id., *Annales bergsonniennes V : Bergson et la politique de Jaurès à aujourd'hui*, P.U.F., Paris 2012; Id., *Prèface*, in C. Zanfi, *Bergson et la philosophie allemande, 1907-1932*, Armand Colin, Paris 2013, tr. it. *Bergson e la filosofia tedesca: 1907-1932*, Quodlibet, Macerata 2013; F. Worms - C. Zanfi, *Bergsoniana*, I, Société des amis de Bergson, 2021 e *Bergsoniana II*, Société des amis

1. Conferma o smentita?

Se un aforisma, una semplice frase o pensiero di per sé autosufficiente possa contribuire alla riapertura e al perfezionamento del cantiere degli studi bergsoniani è la domanda di fronte alla quale sceglie di porsi l'autore. A stupirlo, per alcuni versi giustamente, è che questa frase parli di morte, che si connoti, quindi, come viene rilevato, in quanto «une des rares phrases, chez ce philosophe de la vie, qui portent sur une question pourtant centrale: la mort»[3]. Ma, insieme con lo stupore, questa emersione inaspettata porta con sé anche una conferma, quella della correttezza e legittimità di una lettura di Bergson che Worms ha portato avanti in maniera coerente a partire da uno dei suoi principali studi dedicati al filosofo, *Bergson ou les deux sens de la vie*, pubblicato nel 2004, che, dopo un ventennio esatto, può forse ritrovare una nuova attualità, confermato nella sua coerenza e nel suo rigore scientifico[4]. Ponendosi di fronte a questa frase, quindi, e scegliendo di rinnovare il proprio impegno nei riguardi di un autore che lo ha impegnato per moltissimi anni, lo studioso rimette in questione il proprio lavoro. Ritornandovi sopra alla luce dell'ultima scoperta, egli apre, quindi, un vero e proprio momento di verifica. Di qui la necessità, come sottolinea lo stesso Worms, di «répondre à la première personne du singulier que cet aphorisme résume aussi le coeur d'une discussion voire, toutes choses égales par ailleurs, d'une confrontation "avec Bergson", philosophique bien sûr!»[5]. Il confronto con questa breve frase porterebbe, secondo l'autore, alla conferma della propria lettura del pensiero del filosofo francese. Emergerebbero, infatti, come ben messo in evidenza da Worms nel suo studio del 2004, due sensi della vita[6]. Anzitutto esiste un vincolo, un bisogno, vitale o mortale, uno stimolo, il *primum vivere*, il sopravvivere, *le sur-vivre*, ma questo non basta. C'è anche l'atto interiore, che ci permette di ricrearci, di riproporci nel tempo come slancio, non semplicemente come creature viventi biologicamente. Ecco che, allora, «l'aphorisme inconnu confirme toute cela. On est tourné d'abord vers la limite de la vie, et la mort nous obsède»[7].

L'importante avvenimento storico-filosofico che sta alla base dell'ultimo studio wormsiano, però, non si limita a una conferma della lettura dello studioso, ma ap-

---

de Bergson, 2022. Per l'elenco completo e aggiornato degli studi wormsiani a proposito di Bergson, si veda Id., *Parcours bibliographique*, cit.

3 - F. Worms, *Avec Bergson*, cit., p. 8.

4 - Id., *Bergson ou les deux sens de la vie*, cit.

5 - Id., *Avec Bergson*, cit., p. 9.

6 - Per l'approfondimento degli elementi e delle tematiche qui appena accennati, si rimanda a Id., *Bergson ou les deux sens de la vie*, cit.

7 - Id., *Avec Bergson*, cit., p. 11.

porta altresì una novità. Dal momento che ci rapportiamo a uno dei due sensi della vita tramite la passione, l'amore verso di essa, vale a dire l'attaccamento che proviamo per la nostra stessa esistenza, proprio per questa ragione, amandola, diventiamo indifferenti alla morte. Esiste in altre parole una saggezza umana che, confrontandosi con la vita, riesce a diminuire il peso specifico della morte, a confinarla nell'oblio. Amare la vita e amarla secondo i dettami della filosofia bergsoniana, che significa comprenderne la forza metafisica, il tempo come creatore e prendere atto della sua intrinseca dimensione di durata, vuol dire dimenticarsi del fatto che essa dovrà finire, che avrà un termine.

Questo il punto di partenza, l'approccio, per una postura che non si limiti alla semplice accettazione di una simile caratteristica, ma che allo stesso tempo rilanci, aprendo anche a una discussione che sia appunto *avec*, *con* Bergson. Perché solamente dal dialogo con gli autori, dall'approccio al testo che metta in evidenza la lettera stessa, facendola risuonare, è possibile ricavare una pratica della storia della filosofia – e della filosofia *tout court* – che sia realmente viva e che si possa costituire come risorsa per l'umano.

2. L'umano tra gioia e sofferenza

Lo sguardo di Worms, tuttavia, non si limita all'esplicitazione di quanto di buono e innovativo quest'aforisma potrebbe aver apportato, ma propone altresì l'apertura di alcuni snodi problematici. Il principale ha a che fare con una questione precisa: rendendoci indifferenti alla morte, non smettiamo certo di avervi a che fare durante la nostra esistenza umana. Non è certo la morte nella sua veste astratta a rappresentare quanto c'è di *peggiore* al mondo, ma piuttosto la sua veste concreta, il male, la sofferenza. La via d'uscita rispetto a quest'impasse è indicata dall'autore in questi termini, proponendo la necessità di

> reconnaître que l'on trouve, dans la lutte contre toutes les formes de la mort, la ressource de toutes les joies, tout ce qui donne un sens à la vie, qui ne se réduit certes pas non plus à l'opposition à la mort en général et comport les expériences absolues du meilleur et du bien, et de l'amour. De fait, on ne peut pas parler du pire, comme il faut continuer à le faire, s'il n'y a pas le meilleur. En ce sens, cet aphorisme comportera toujours une orientation, pour aller plus loin[8].

Rimanendo entro il recinto del pensiero bergsoniano e della sua letteratura critica, non ci si può porre di fronte al problema del male, della sua realtà e consistenza per l'esistenza umana, senza interrogare *Le due fonti della morale e della reli-*

8 - *Ibid.*, p. 14.

gione[9]. Il male, così come la sofferenza che ne emerge, non possono essere negati: la sofferenza, l'esistenza del male, vengono presi sul serio dal filosofo francese, che non può lasciare da parte la questione dell'utilità del pensiero[10]. Come si pone il filosofo di fronte all'inesauribile interrogativo sul senso dell'esistenza umana e sulla sua sofferenza? Un buon atteggiamento, a questo proposito, sembra quello di andare oltre, tramite la sottolineatura di un'opposizione della realtà della sofferenza, concreta, che ha strettamente a che fare con la realtà degli esseri umani, al problema del male, fittizio, astratto, che non coinvolge davvero l'essere umano. Fare questo significa, allora, concentrarsi su gioia e sofferenza, non su bene o male. Il problema, tuttavia, pare rimanere intatto: come può la gioia combattere, bilanciare concretamente la sofferenza? Se è vero che il mistico, in Bergson, accede a una gioia definitiva, come può questo stato di gioia fare da contrappeso al dolore, per esempio, di una madre in lutto per la perdita di un figlio?

La possibile soluzione, che Worms si sente di tracciare come eventualità forse praticabile, consiste nell'abbandono di una logica di calcolo e nell'apertura di un orizzonte speculativo diverso. Il problema della sofferenza è un problema reale e porsi di fronte a una simile questione implica, come passo preliminare, un fondamentale discernimento tra quelli che sono i problemi veri, reali e quelli che non lo sono[11]. Ecco che, allora, bisogna necessariamente ammettere l'esistenza umana come giocata tra due assoluti, gioia e sofferenza: un'esistenza *tra*, *entre*, poiché «Le problème de la souffrance et de la joie ne revient donc pas à rélativiser l'une par l'autre, mais au contraire à constater le déchirement de notre vie, de la vie, entre deux absolus»[12]. Non rimane, quindi, solamente lo scandalo del male a far inorridire l'uomo, a metterlo di fronte alla sua sostanziale incapacità di agire e di re-agire di fronte allo scacco della propria esistenza, ma anche il ben più enigmatico scandalo della gioia. Si apre, in questo modo a una «vie morale concrète de celui qui posera, réellement, la question de la souffrance, du crime, en lui comme en dehors de lui, et qui s'étonnera aussi, réellement, de la force de joie ou de la res-

9 - H. Bergson, *Le due fonti della morale e della religione*, Laterza, Roma-Bari 2020, ed. fr. Id., *Les Deux Sources de la morale et de la religion*, édition critique par F. Keck et G. Waterlot, P.U.F., Paris 2008.

10 - F. Worms, *"Terrible réalité" ou "faux problème"? Le mal selon Bergson*, in Id., *Avec Bergson*, cit., pp. 25-37. Riferimenti privilegiati di Worms in queste pagine sono sia l'interpretazione, ormai classica, che di Bergson diede Vladimir Jankélévitch, che gli studi di Jean Nabert. Cfr. V. Jankélévitch, *Henri Bergson*, P.U.F., Paris 2015; J. Nabert, *Élements pour une éthique*, Aubier-Montaigne, Paris 1993[3] (P.U.F., Paris 1943) e Id., *Essai sur le mal*, Aubier-Montaigne, Paris 1970 (P.U.F., Paris 1955).

11 - «Ce qui importe surtout, nous semble-t-il, c'est ceci: la souffrance n'est pas un faux mais un vrai problème, auquel le mystique apporte une réponse non pas seulement théorique mais pratique» (F. Worms, *Avec Bergson*, cit., p. 32).

12 - *Ibid.*, p. 33.

source de bonté et de justice qui ne cessent de s'y opposer»[13]. Una caratterizzazione che rimanda all'apertura altresì di una condizione umana che necessita di uno spazio di libertà. Quale libertà? In che orizzonte concepirla? «il faut bien aussi *qu'entre ces deux réalités* s'ouvre l'espace de la liberté, de la conscience, de la responsabilité, de la faute méditée, du bien voulu – bref, de ce que Bergson appelle, à raison, la "*vie morale*"»[14].

3. Un'apertura: l'umano tra apertura e chiusura verso la tecnologia?

Uno sguardo d'insieme su questo testo wormsiano, quindi, può essere gettato a partire dalla considerazione dell'*incipit* del quarto e ultimo capitolo de *Le due fonti*. Se comprendiamo l'operazione bergsoniana di articolazione di un orizzonte in cui apertura e chiusura siano compresenti, dobbiamo anche tenere a mente – come suggerisce di fare Worms – che tale distinzione può conservare senza dubbio la sua cogenza teoretica, non senza però sacrificare qualcosa sul piano pratico. Lo sguardo deve tornare a posarsi su un'umanità che si trova, ora più che in altre epoche, dilaniata *entre* apertura e chiusura, in misura ancora maggiore in quanto tale umanità si trova continuamente a relazionarsi con la meccanica, con le macchine, con la tecnologia e le sue frontiere. Un'umanità il cui soggetto non può che essere concepito in quanto *homo faber*. La posta in gioco, in altre parole, è la seguente: l'essere umano è giocato *tra* apertura e chiusura, tra società e morale aperte e chiuse, tra religione dinamica e religione statica, se seguiamo Bergson, anche e soprattutto perché è in costante sfida, in continuo dialogo, con le istanze del progresso tecnologico. Come guardare, quindi, a un autore di questa rilevanza? L'indicazione di Worms è chiara:

> Tel serait donc le sens de cette utilité pratique pour "nous" et "maintenant". Le clos et l'ouvert peuvent et doivent aider dans l'orientation que l'on donnera à cette technique moderne, surpuissante, qui peut soit conduire à des destructions inédites du côté de la clôture, soit se tourner plus fortement que jamais du côté de l'ouverture[15].

E ancora:

> La liberté de l'humanité, "maintenant", exige cette nouvelle alliance du mécanique et du mystique, par-delà celle, nocturne, deux fois close, de la guerre,

13 - *Ibid.*, p. 37.

14 - *Ibid.*, p. 36, il corsivo è dell'autore.

15 - *Ibid.*, p. 45.

pour laquelle la force technique se met au service non pas du mystique, mais du mythique, du mythologique, de l'idéologique, et de la "fabulation"[16].

FRANCESCO PATRONE
PhD in Scienze Umane, curriculum filosofico, Università degli Studi di Perugia - docteur canonique en Philosophie, Université Catholique de Lyon
francesco.patrone@dottorandi.unipg.it

16 - Ibidem.



Sapienza di Dio che scende dal Cielo
e scienze dell'uomo che salgono dalla Terra
diventano "uno" al centro:
la mente di Gesù (1 Cor 2,16)
a tutti partecipata nel dono dello Spirito Santo


Istituto Universitario
SOPHIA
University Institute