



Filosofia Morale / Moral Philosophy

vol. 2, 2024

FILOSOFIA MORALE/MORAL PHILOSOPHY

La rivista della Società Italiana di Filosofia Morale



Direttore Responsabile

Adriano Fabris

Direzione Editoriale

Anna Donise, Roberto Mordacci

Redazione

Fabrizia Abbate, Mariafilomena Anzalone, Giulia Battistoni, Luca Bertolino, Stefano Biancu, Calogero Caltagirone, Alessandro Chiessi, Carlo Chiurco, Giovanni Cogliandro, Giovanna Costanzo, Michel Croce, Marco Deodati, Roberto Formisano, Simone Grigoletto, Stefania Lombardi, Fabio Mazzocchio, Claudia Navarini, Orietta Ombrosi, Annia Pia Ruoppo, Maria Russo, Simone Stancampiano

Comitato scientifico

Stefania Achella, Francesco Adorno, José Ángel Agejas Esteban, Andrea Aguti, Luigi Alici, Robert Audi, Roberto Celada Ballanti, Maria Cristina Bartolomei, Rossella Bonito Oliva, Angelo Campodonico, Giovanna Costanzo, Antonio Da Re, Mario De Caro, Adriano Fabris, Jean Marc Ferry, Luca Fonnesu, Rainer Forst, Marina Garcés Mascareñas, Benedetta Giovanola, Paolo Gomasasca, Vittorio Höslé, Irene Kajon, Sandro Mancini, Carmelo Meazza, Francesco Miano, Alessio Musio, Julian Nida-Rümelin, Riccardo Panattoni, Maria Teresa Pansera, Corinne Pelluchon, Adriano Pessina, Vallori Rasini, Paola Ricci Sindoni, Alfredo Rocha de la Torre, Maria Vita Romeo, Laura Santos, Giovanni Scarafile, Sally Sedgwick, Petr Urban, Maria Silvia Vaccarezza, Klaus Vieweg, Carmelo Vigna, Marcello Vitali Rosati, Linda Zagzebski

Tutti i saggi scientifici vengono sottoposti a double-blind peer review

MIMESIS EDIZIONI (Milano – Udine)

www.mimesisedizioni.it

<https://mimesisjournals.com/ojs/index.php/MF/index>

ISSN: 2785-5457

ISBN: 9791222318318

© 2024 – MIM EDIZIONI SRL

Piazza Don Enrico Mapelli, 75

20099 Sesto San Giovanni (MI)

Phone: +39 02 24861657 / 21100089

Indice

5 Anna Donise, Roberto Mordacci, *Editoriale*

Articoli

- 15 Marco Damonte, "Dependency worker": *ruolo e valore di una nozione a partire da Eva Feder Kittay*
- 37 Maria Teresa Pacilè, *Per una rinascita spirituale dell'uomo. Karl Jaspers, la svolta assiale e il progetto della Weltphilosophie*
- 61 Lars Rensmann, *Antinomies of Progress. Notes on Adorno's Critical Theory and the Concept of Progress in Our Time*
- 81 Michele Vadilonga Gattermayer, *From the Anthropological Machine to the Animal Crisis*
- 101 Ivano Zanzarella, *The Problem of Musical Creativity and its Relevance for Ethical and Legal Decisions towards Musical AI*

Discussioni

Intelligenza Artificiale: sfide e controversie.

A cura di Maria Filomena Anzalone e Maria Russo

- 119 Giuseppe De Ruvo, *Intelligenza artificiale, sicurezza nazionale e capitalismo politico. La crisi della riflessività democratica e la necessità di un'etica strategica*
- 131 Eran Fisher, *AI as epistemic media and the future of subjectivity*
- 143 Fabio Fossa, *Autonomia, automazione, mediazione. Alcune note critiche sull'etica dei veicoli a guida autonoma*
- 155 Pasquale Grieco, *Design for action: sviluppo tecnologico e la metamorfosi delle categorie etiche*
- 173 Giorgia Redi, *RoboCare. Anziani più indipendenti o più invisibili?*
- 189 Simona Tiribelli, *Artificial Intelligence and Moral Progress: A Missing Debate*

Recensioni

- 203 Nunzio Bombaci, *La persona, il prossimo, l'amico. Le figure dell'altro*, in Pedro Laín Entralgo, ETS, Pisa 2023, pp. 296 (Bryan Jesus Irias Alfaro)
- 207 Francesco Paolo Adorno, *The Transhumanist Movement*, Palgrave Macmillan, Cham 2021, pp. 233 (Roberto Di Ceglie)
- 213 Ortensio Zecchino, *Perché non possiamo non dirci "cristiani". Letture e dispute sul celebre saggio di Benedetto Croce*, Rubbettino Editore, Saveria Mannelli 2024, pp. 254 (Michele Farisco)
- 219 Roberto Mordacci, *Critica e Utopia. Da Kant a Francoforte*, Castelvecchi, Roma 2023, pp. 185 (Adriano Lotito)
- 227 Simona Tiribelli, *Identità personale e algoritmi. Una questione di filosofia morale*, Carocci, Roma 2023, pp. 131 (Angelo Tumminelli)

Anna Donise* e Roberto Mordacci**

Editoriale

In un romanzo del 2019, *Machine like me*, Ian McEwan mette in scena il classico conflitto tra essere umano e robot, raccontando le vicende dell'androide Adam, una macchina perfetta, con le sembianze di un bell'uomo moro, con la pelle morbida e calda, capace di comporre poesie e di assumere diverse espressioni facciali. Ma la vera particolarità di questo androide è quella di essere dotato di un software che – attraverso il confronto con migliaia o forse milioni di dilemmi etici – lo rende capace di diventare una macchina morale. McEwan, con una buona dose di ironia, immagina un mondo in cui lo sviluppo dell'IA prende una direzione interessante o forse spaventosa: mentre le auto a guida autonoma non riescono a superare la prova del traffico e dei pirati informatici, la tecnologia si spinge fino a realizzare il sogno di “una virtù robotica in grado di redimerci”; l'oggetto della redenzione siamo noi umani, noi esseri imperfetti, anzi “moralmente difettosi: incoerenti, labili a livello emotivo, inclini al preconetto, all'errore di valutazione, spesso per ragioni di egoismo”.

Il tema della sezione *Discussioni* di questo numero di FM/MP – *Intelligenza Artificiale: sfide e controversie*, egregiamente curato da Maria Filomena Anzalone e Maria Russo – in linea con la riflessione dello scrittore inglese, pone il problema del rapporto tra umano e IA, provando a portare argomenti utili al dibattito etico. È chiaro che quando la riflessione filosofica si confronta con la IA il tema che emerge e diviene centrale è quasi sempre l'umano con le sue trasformazioni reali o solo possibili, auspicabili o temibili. Nel contributo che apre la sezione, Eran Fisher si chiede proprio che cosa significhi essere umani in un mondo sempre più popolato dall'IA, proponendo di pensare queste nuove tecnologie nel contesto di una storia molto più lunga, non digitale: quella dei “media epistemic”, mezzi capaci di ampliare le capacità cognitive degli utenti umani.

La connessione tra IA e possibile sviluppo morale umano è centrale sia nel contributo di Fabio Fossa, dedicato al problema delle macchine

* Università degli Studi di Napoli Federico II – anna.donise@unina.it.

** Università Vita-Salute San Raffaele, Milano – mordacci.roberto@univr.it.

a guida autonoma, sia nel contributo di Simona Tiribelli, che discute di progresso morale. Nel contributo di Fossa viene messo in evidenza come la desiderabilità delle tecnologie di IA si trova ad essere giustificata anche facendo riferimento alla loro supposta capacità di apportare benefici etici, ovvero alla loro capacità di migliorare condizioni che attualmente, affidate esclusivamente all'umano, risultano essere moralmente insoddisfacenti. Al centro della riflessione di Fossa c'è il valore dell'autonomia e l'obiettivo del contributo è indagare l'ambiguità con la quale viene affrontato questo tema in relazione ai veicoli automatizzati e più in generale nelle tecnologie legate all'IA. Tiribelli affronta invece l'impatto della tecnologia dell'IA sul progresso morale a livello collettivo, e argomenta a favore della rilevanza di una progettazione di sistemi di IA in grado di proteggere le nostre conquiste morali e di implementare il progresso morale della società. Se, in effetti, tali tecnologie mostrano, in linea teorica, il potenziale di influire negativamente, le stesse, se riprogettate, hanno il potenziale per diventare forze di promozione concreta del progresso morale. La progettazione dello sviluppo dell'IA è il filo conduttore anche nel saggio di Pasquale Grieco che si concentra ancora sui valori umani e sulle loro trasformazioni. Grieco inquadra la questione alla luce del *design turn* e nella sua prospettiva le trasformazioni più interessanti e significative riguardano la soggettività, la mediazione tecnologica e la *agency*. Il problema della regolamentazione dell'IA e delle questioni etiche poste dalle nuove tecnologie attraversa anche il contributo di Giuseppe De Ruvo, che mette in evidenza la connessione problematica tra governance dell'IA e le burocrazie della difesa e della sicurezza degli stati. Indagando la relazione tra gli apparati di sicurezza nazionale, il mercato e le pratiche deliberative pubbliche, la riflessione filosofica sull'IA deve tematizzare a fondo il mutevole rapporto tra Stato e mercato, non tralasciando neanche le questioni geopolitiche.

Un tema utile sia a mettere a fuoco alcune delle trasformazioni dell'umano che derivano dalla presenza massiva delle nuove tecnologie nelle nostre vite, sia a porre il problema della relazione, tesa ma rilevante, tra agire moralmente connotato e IA, è il ruolo dei robot all'interno della società. Di particolare interesse è il rapporto con i *Socially Assistive Robots* (SAR), oggetto del contributo di Giorgia Redi. Questi robot sono progettati per alleviare la solitudine e fornire aiuto nelle attività di cura e assistenza agli anziani ed è evidente che il rapporto con essi obbliga a ripensare il modo di concepire e vivere le relazioni.

La sezione Articoli, come di consueto, non ha un tema unico, ma raccoglie contributi su temi diversi. L'articolo di Marco Damonte è dedicato a un'analisi delle tesi di Eva Feder Kittay sul tema della vulnerabilità e della dipendenza, elaborate soprattutto nella raccolta di saggi *La cura*

dell'amore. Donne, uguaglianza, dipendenza del 1999 (pubblicato in Italia nel 2010). La figura chiave di Kittay è il *dependency worker*, ossia colei o colui che “presta aiuto a una persona incapace di provvedere a se stessa”: questo “lavoro di dipendenza” è un aspetto costitutivo dell'umano, che è tuttavia sconosciuto dall'antropologia presupposta dalle teorie liberali. La relazione di dipendenza non è propria solo dei minori o delle persone con disabilità: è una condizione che ogni essere umano attraversa in varie fasi della propria esistenza. Tenerne conto nelle teorie della giustizia è essenziale, perché queste – per non essere riduttive e discriminatorie – devono prevedere l'esplicito riconoscimento e sostegno delle relazioni di servizio (espresse dall'idea greca di *doulia*).

La condizione della natalità, esplorata da Hannah Arendt e legata alla dipendenza, è fra i moventi della riflessione di Karl Jaspers sull'età assiale, oggetto del contributo di Maria Teresa Pacilè. Per Jaspers, l'età assiale costituisce un momento di (ri-)nascita spirituale dell'umanità autentica: emerge in quel non breve volgere storico (fra il IX e il III secolo a.C.) una “metamorfosi totale della storia”, ossia una trasformazione dell'esperienza umana che attraversa tutte le civiltà e che segna il sorgere delle culture “classiche” di ogni tradizione. Jaspers caratterizza questo periodo come una “nascita” spirituale, ovvero come il sorgere di una “coscienza” in cui si formano le categorie fondamentali dell'esperienza. Pacilè analizza le numerose critiche e riprese (fra cui quelle di Assman e di Eisenstadt) di questa concezione della filosofia della storia, mostrando la vitalità dell'intuizione jaspersiana in un momento in cui la storia si fa di nuovo questione problematica e moralmente impegnativa.

La filosofia della storia sta in effetti vivendo una rinnovata stagione e uno dei dibattiti tornati al centro dell'interesse è quello intorno al tema del progresso. Il contributo di Lars Rensmann esplora la nozione di progresso (morale e sociale) a partire dalla riflessione di Theodor W. Adorno, mettendone in luce l'attualità e la valenza critica. In un'“età catastrofica” come la attuale, risulta estremamente problematico articolare la relazione – necessaria – fra progresso e “vita buona”: piuttosto, la celebrazione acritica del progresso in termini tecnologico-scientifici occulta sistematicamente le forme di regresso sociale e morale che si producono nella società contemporanea. Le spinte autoritarie che permeano le aspirazioni ipertecnologiche e transumaniste di persone e gruppi molto influenti nel quadro attuale mostrano le contraddizioni letali cui è esposta una concezione riduttiva del progresso. A questa deriva, Rensmann oppone la necessità di una riformulazione della solidarietà cosmopolita, della democrazia e della critica.

Il confine tra umano e animale, e dunque l'interrogativo antropologico, è al centro del contributo di Michele Vasilonga Gattermayer, che discute criticamente l'idea di “macchina antropologica” proposta da

Giorgio Agamben. Come diversi studiosi dell'etica animalista hanno sottolineato, la concezione di Agamben esclude sistematicamente dallo sguardo il lato animale della "macchina", relegando i viventi non-umani in una zona d'ombra. Questa esclusione è alla radice della "crisi animale" che si concretizza non solo nel trattamento violento, oppressivo e brutale degli animali, ma più radicalmente in una deriva ontologico-antropologica in cui la figura del dominio risulta la cifra caratteristica della civiltà umana.

Un'attività che appare tipicamente umana, ma che oggi è progressivamente "colonizzata" dalle tecnologie, e in particolare dall'intelligenza artificiale, è la musica. L'articolo di Ivano Zanzarella illustra i molti modi in cui la creatività musicale è oggi non semplicemente imitata dall'IA ma anche ricreata e sostenuta, in un utilizzo sempre più ampio delle tecnologie nella produzione di musica. Le conseguenze della Musical Artificial Intelligence (MAI) non riguardano solo l'estetica ma anche, se non principalmente, l'etica: non si tratta di "arginare" l'invasione della creatività da parte dell'IA, che è già ampiamente utilizzata dai compositori, bensì di regolare il suo impiego a partire dalla presa d'atto delle sue capacità creative e originali. L'attribuzione di uno status morale ai sistemi artificiali di produzione musicale è, argomenta Zanzarella, la premessa necessaria per affrontare nel merito la questione della creatività, senza assumere aprioristicamente che essa sia esclusivo appannaggio dell'umano.

Complessivamente, gli articoli mostrano che la ricerca sull'umano, la sua fragilità, la sua natura storica, il suo rapporto da un lato con la vita animale e dall'altro con le tecnologie, è sfidata profondamente nel contesto contemporaneo. La comprensione di chi o che cosa siamo, e di quale sia la "posizione dell'umano nel cosmo" torna a imporsi fortemente in questi primi venti anni del millennio, così come accadde a inizio Novecento. Si tratta di questioni eminentemente morali, cui l'indagine scientifica, tecnologica e anche solo teoretica non sembra offrire le risposte necessarie. In questo senso, questo fascicolo testimonia non solo della vitalità della filosofia morale ma anche della sua capacità di intercettare le questioni urgenti del nostro tempo.

La rubrica delle *Recensioni* è, anche in questo numero ricca e vivace, mostrando l'ampiezza della ricerca relativa ai temi morali. Invitiamo tutti gli interessati, autori e recensori, a segnalarci lavori da censire e da presentare attraverso la rivista.

La risposta alla call for papers della sezione *Discussioni* di questo numero è stata particolarmente intensa, forse perché il tema prescelto, oltre a essere molto presente nel dibattito pubblico, è vissuto come attuale e urgente. Questo, se da un lato è indice dello stato di salute della rivista, che comincia a diventare punto di riferimento della comunità di filosofi morali, dall'altro ci ha obbligati a una selezione più drastica di quanto

accaduto per i numeri passati e la percentuale di testi accolti dopo la valutazione in doppio cieco da parte dei revisori è stata particolarmente bassa. Il lavoro della Redazione per queste operazioni è impegnativo e un ringraziamento va a tutti coloro che ne fanno parte, nonché ai revisori che generosamente accolgono la richiesta di valutare i contributi.

La Redazione ha definito i temi per i prossimi due numeri della sezione *Discussioni*. Per il fascicolo 1/2025 il tema sarà **Giustizia sociale e nuove tecnologie: sfide etiche e opportunità**. Per il fascicolo 2/2025 il tema sarà **Cinema, cittadinanza ed educazione: plasmare immaginari democratici**

Riportiamo qui le rispettive call for papers.

1/2025: Giustizia sociale e nuove tecnologie: sfide etiche e opportunità

La call for papers è rintracciabile sul sito della rivista e la scadenza per l'invio dei contributi (compresi fra 15.000 e 30.000 caratteri spazi inclusi) è il 20 febbraio 2025. I contributi vanno caricati sul portale della rivista.

L'argomento è affrontato da una prospettiva etica e può essere trattato con riferimento alle seguenti questioni:

1. Sollevare la questione di una giustizia sociale più equa richiede una valutazione costante della ricchezza e della sua distribuzione, della possibilità di pari opportunità, della difesa di diritti che garantiscono il rispetto della dignità di ognuno, della presenza di misure compensative e di sostegno dei soggetti più vulnerabili. Considerare, inoltre, il ruolo rivestito nella sfera sociale dalle istituzioni in grado di garantire servizi pubblici, come la sanità e la scuola pubblica, il diritto al lavoro, la previdenza sociale, il sistema di tassazione, consente non solo di comprendere diritti e doveri dei cittadini, ma di valutare quanto siano centrali in un determinato contesto valori come uguaglianza e chance di vita in un tempo in cui le differenze si accrescono e in cui gli algoritmi sembrano decidere per i più nella lotteria delle opportunità.
2. L'impatto delle nuove tecnologie (fra cui l'IA) nelle attuali società, di fatto, complica le questioni etiche, poiché bisogna chiedersi quanto un uso "responsabile" della tecnologia sappia promuovere maggiore equità, giustizia, sicurezza. Una equità nelle cure, una equa distribuzione delle risorse, pari opportunità per fasce diverse della popolazione, la consapevolezza di essere trattati con "giustizia" e che siano rispettate le legittime richieste per una uguaglianza di genere, economica, etnica.
3. Affrontare i temi della giustizia sociale richiede la necessità di affrontare in maniera significativa i temi della uguaglianza e della

pari opportunità: come è possibile una società più equa alla luce delle imprescindibili differenze fra cittadini?

4. Nel concetto di giustizia sociale è possibile includere nuovi paradigmi come quelli di giustizia riparativa, nei confronti per esempio dei detenuti? Come promuovere percorsi di inclusione per i migranti e per le persone con disagio psichico?
5. Le nuove tecnologie possono contribuire ad una pari opportunità nell'accesso alle cure e ad un sistema sanitario più equo? Ad un sostegno nei confronti dei disabili?
6. Grazie alla AI sembra esserci un allargamento della sfera della uguaglianza, in ordine alle possibilità aperte ai più di accesso alle informazioni, ma come valutare il fatto che questo avviene attraverso una diminuzione della privacy e di un sistema di sorveglianza continua?
7. Le tecnologie informatiche in ambito sanitario possono promuovere una più equa giustizia sociale diminuendo i gap generazionali, di genere, sociali, economici? Come si può promuovere un miglioramento nelle condizioni di vita di anziani e di persone con disabilità?

2//2025: Cinema, cittadinanza ed educazione: plasmare immaginari democratici

La call for papers è rintracciabile sul sito della rivista e la scadenza per l'invio dei contributi (compresi fra 15.000 e 30.000 caratteri spazi inclusi) è il 20 settembre 2025. I contributi vanno caricati sul portale della rivista.

Il fascicolo esplora il ruolo del cinema nella costruzione di comunità più giuste, inclusive e democratiche con particolare attenzione al cinema come strumento educativo in grado di favorire il pensiero critico, di sviluppare l'immaginazione utopica e di rafforzare il senso di appartenenza civica.

L'argomento può essere sviluppato rispondendo alle seguenti domande:

1. In che modo il cinema, con il suo linguaggio narrativo e visivo unico, può trasformarsi da mezzo di intrattenimento a strumento per l'educazione filosofica, presentando e argomentando posizioni speculative e morali?
2. Quali metodi innovativi possono essere sviluppati per integrare il cinema nelle pratiche educative, veicolando esperienze partecipative, stimolando un dialogo che non si limiti a pensare alle tesi esplicitamente esposte nei film, favorendo il pensiero critico e la riflessione etica?
3. In che modo l'immaginazione di futuri migliori o peggiori, per esempio grazie alla raffigurazione di scenari utopici e dispotici, nei film può aiutarci ad analizzare criticamente e affrontare le contraddizioni e le sfide del presente?

4. Come le speculative proposte dai registi possono ispirare il pubblico ad affrontare questioni come le disuguaglianze, il degrado ambientale, il progresso tecnologico e le crisi politiche? Come può la rappresentazione di società alternative nella finzione cinematografica generare un'azione trasformativa nel mondo reale?

5. È possibile che il cinema, con il suo potere unico di unire le persone, alimenti un senso condiviso di responsabilità all'interno delle società democratiche e contribuisca a colmare divisioni, affrontare polarizzazioni e promuovere la partecipazione civica?

6. Il modo in cui i film affrontano le questioni di diversità, equità e inclusione hanno davvero un impatto sulla percezione culturale e sulle politiche pubbliche? In che modo i film influenzano le conversazioni sulla ricerca dell'identità?

7. I film possono coinvolgere emotivamente gli spettatori, spingerli a confrontarsi con dilemmi morali e a mettere in discussione norme sociali e valori consolidati. Fino a che punto un film riesce ad alimentare l'empatia e a sviluppare l'immaginazione etica?

8. La narrazione cinematografica si basa spesso sull'immedesimazione dello spettatore in personaggi e circostanze specifiche, inducendo il pubblico a problematizzare le proprie idee su che cosa è bene e che cosa è male: quali sono potenzialità e limiti delle strategie narrative d'immedesimazione?

Articoli

Marco Damonte*

“Dependency worker”: ruolo e valore di una nozione a partire da Eva Feder Kittay

Abstract

“Dependency worker”: role and value of a notion starting with Eva Feder Kittay

More than twenty-five years after the publication of Kittay’s *Love’s Labor. Essays on Women, Equality, and Dependency*, it is desirable to reread this text through the notion of dependency worker. With reference to this concept, it is possible to grasp some original aspects of Kittay’s reflection neglected by secondary literature that go beyond the classic setting of an ethic of care, but which prove to be useful in providing a contribution both to feminist philosophy and to disability studies. After an adequate contextualization of Kittay’s book and some clarifications on the style in which the arguments are presented, I will identify the starting point of Kittay’s reflections in the anthropological notion of vulnerability, understood in an ontological dimension. Precisely because of the vulnerability that constitutes human beings, each person must be considered “some mother’s child” and treated accordingly. Thanks to the notion of vulnerability, the dependency work assumes a particular importance which I will highlight by providing an adequate definition, a basic classification and indicating its main characteristics. Considering the current conceptions of justice and equality inadequate to the needs of dependency work, Kittay proposes to abandon the current welfare model and move to a model based on “*doulia*”. I will conclude by suggesting the usefulness of Kittay’s reflection on dependency work to appreciate and enhance the legislative proposal for the protection of the “family caregiver” currently under discussion in the Italian parliament.

Keywords

Eva Feder Kittay; Dependency Work; Transparent Self; Vulnerability; *Doulia*

* Università degli Studi di Genova.



1. *La cura dell'amore: testo e contesto*

A quasi un quarto di secolo dalla sua prima pubblicazione nel 1999 e a poco più di dieci dalla sua traduzione italiana nel 2010, *La cura dell'amore. Donne, uguaglianza, dipendenza* di Eva Feder Kittay conserva ancora la sua attualità. Questo arco di tempo, per il metro della filosofia contemporanea, è troppo breve per stabilire se quest'opera sia destinata a diventare un classico, ma sufficiente per constatarne l'originalità e metterla in dialogo con le più recenti proposte di etiche del *care*¹. In questo ambito, le sue potenzialità sono state solo in parte sfruttate a causa di una intrinseca complessità del testo dovuta alla sua genesi, alla molteplicità dei temi affrontati e alla metodologia utilizzata per la sua stesura. Al fine di renderlo fruibile nel dibattito sul *care* propongo quale chiave di lettura quella del *dependency worker* spesso trascurata nella letteratura secondaria perché relegata a una dimensione di eccezionalità, quale quella della cura prestata nei confronti di persone affette da gravi disabilità incapaci di far fronte ai propri bisogni primari².

Questa categoria, poco tematizzata dalla riflessione filosofica, se non nel caso più ristretto del *caregiver*³, interessa trasversalmente i numerosi settori che hanno impegnato Kittay nel corso del suo itinerario intellettuale: dalla filosofia femminista di matrice americana, ai *disability studies*, dall'etica della cura, fino alla teoria della giustizia nella sua duplice accezione politica e antropologica. Tutti questi ambiti si ritrovano in *La cura dell'amore* che, infatti, si presenta come un testo composito, di fatto una raccolta di materiali pubblicati nel corso dell'ultimo decennio del secolo scorso. Sebbene l'autrice li abbia rivisti al fine di renderli omogenei, non è immediato individuare il perno della mutazione delle forme dialettiche intorno al quale Kittay fa dipendere la sua proposta. L'argomento che intende sostenere è precisato in questi termini: «la tesi del mio libro è che la nostra dipendenza reciproca non può essere messa tra parentesi senza escludere parti significativa della nostra vita e della popolazione dalla sfera dell'uguaglianza»⁴. Tale tesi si staglia però in un orizzonte più ampio, costituito da un interrogativo che guida un più vasto filone di ricerca così formulato: «perché l'uguaglianza si è dimostrata così irraggiungibile per le donne?»⁵. Considerare la tesi di *La cura dell'amore* come una mera tappa di questo filone sarebbe

¹ Cfr. F.P. Adorno, *Gli obblighi della cura. Problemi e prospettive delle etiche del care*, Vita e Pensiero, Milano 2019.

² Ivi, p. 11.

³ Cfr. O. Giolo, B. Pastore (a cura di), *Vulnerabilità. Analisi multidisciplinare di un concetto*, Carocci, Roma 2018, pp. 244-247.

⁴ E.F. Kittay, *La cura dell'amore. Donne, uguaglianza, dipendenza* [1999], tr. it. a cura di S. Belluzzi, Vita e Pensiero, Milano 2010, pp. XXXIV-XXXV.

⁵ Ivi, p. XVIII.

però riduttivo: tale tesi, proprio nella misura in cui prende in considerazione lo statuto del *dependency worker* e il suo ruolo nella società, si presenta non come risposta a un caso particolare – per quanto più diffuso di quanto si potrebbe pensare –, ma come «propedeutica a ogni futura teoria dell’uguaglianza»⁶. Resta comunque una certa oscillazione tra il tentativo (pragmatico) di affrontare le problematiche connesse al *dependency work* e la proposta (teoretica) di considerarlo emblematico rispetto all’antropologia e fondativo rispetto a un nuovo patto sociale.

Oltre che a livello tematico, anche sul piano metodologico il testo presenta una varietà di approcci non sempre facilmente riconducibili a una trama unitaria. La stessa autrice dichiara come il suo lavoro dipenda solo in minima parte dagli assunti propri della filosofia convenzionale anglo-americana, assunti messi sempre più spesso in discussione⁷. Nonostante ciò, l’orizzonte entro cui Kittay propone la sua riflessione e gli autori con cui si confronta sono riconducibili alla filosofia analitica. Kittay però, pur mantenendo il rigore argomentativo e insistendo sulla chiarezza delle definizioni man mano offerte, evita di ricorrere a formalizzazioni logiche e a definizioni necessarie e sufficienti. Accanto al registro argomentativo, nel testo si incontrano registri stilistici differenti, quali quelli riconducibili alla persuasione, tra cui testimonianze in prima persona – Kittay è madre di Sessa, una giovane donna affetta da una grave disabilità⁸ –, riflessioni tratte da racconti di cronaca ed esempi di funzionamento tipici dello stato sociale americano. L’empatia creata con il lettore grazie all’accortezza di corroborare dati, percentuali e *report* legislativi con aspetti più narrativi, costituisce un appello anche emotivo che non sempre aiuta il lettore, spinto più a solidarizzare con le persone in situazione di vulnerabilità che non a seguire i sottili ragionamenti volti a cambiare lo *status* di chi le prende in carico. Per comprendere la relazione tra *dependency worker* ed etica della cura sulla scia di Kittay si tratta di considerare, in primo luogo, quale sia l’antropologia di riferimento dell’autrice, un’antropologia basata su una concezione peculiare di vulnerabilità.

2. La vulnerabilità umana

A differenza del pionieristico lavoro di Robert Goodin che riflette sulla vulnerabilità potenziale di una persona rispetto alle azioni di terzi⁹, Kittay muove da un dato fondamentale dell’esperienza umana: quello

⁶ Ivi, p. XXXIII.

⁷ Cfr. Ivi, p. XVII.

⁸ Cfr. Ivi, pp. 267-291.

⁹ Cfr. R. Goodin, *Protecting the Vulnerable*, Chicago University Press, Chicago 1985.

della dipendenza, emblematicamente espresso nelle forme di non autosufficienza e simboleggiato dalle disabilità più gravi. Di fronte a questa osservazione diventa lapalissiano che l'autonomia, la libertà, l'autorealizzazione e l'indipendenza non costituiscono lo statuto antropologico dell'essere umano in modo originario: piuttosto sono ideali regolativi a cui troppe antropologie moderne – quelle però che, però, non a caso, informano il nostro modello socio-economico – tendono. Tali caratteristiche ineriscono l'essere umano come potenzialità la cui realizzazione, quando non compromessa, è resa possibile da una rete di relazioni e di dipendenze che la sorreggono e a prescindere dalla quale restano latenti e inesprese. Kittay non si sofferma sugli autori che storicamente hanno consolidato il nesso tra antropologia dell'autonomia¹⁰ e fondazione delle istituzioni politiche, ma un esplicito riferimento a Thomas Hobbes è sufficiente a suggerire come per lei ogni forma di contrattualismo sociale riposi inevitabilmente su una antropologia carente, se non fuorviante¹¹. Purtroppo, Kittay non dimostra grande sensibilità storica, pertanto le sue intuizioni meriterebbero di essere corroborate da studi più puntuali, a cominciare proprio da un approfondimento sulla posizione di Hobbes¹².

In ossequio alla tradizione femminista e al suo impegno nei confronti della disabilità¹³, Kittay fa risiedere l'aspetto della vulnerabilità umana su una adeguata valorizzazione del corpo e della biologia umana¹⁴, ma compie un passo ulteriore che le evita di associare la vulnerabilità alla mera dimensione fisica, affermando:

potremmo anche dire che il lungo processo di maturazione degli uomini combinato con le capacità spiccatamente umane dell'affetto e del senso morale rende il prendersi cura delle persone dipendenti un segno distintivo della nostra umanità. La nostra dipendenza, quindi, non è solo una circostanza eccezionale.¹⁵

In questo modo Kittay riformula i criteri della personalità morale¹⁶ e contribuisce in maniera decisiva alle discussioni correnti circa lo statuto della vulnerabilità, così sintetizzate da Orsetta Giolo e Baldassare Pastore:

¹⁰ Cfr. C. Mackenzie, N. Stoljar (eds.), *Relational Autonomy. Feminist perspectives on Autonomy, Agency, and the Social Self*, Oxford University Press, New York 2000.

¹¹ Cfr. E.F. Kittay, *La cura dell'amore. Donne, uguaglianza, dipendenza*, cit., p. 41.

¹² Cfr. E. Pulcini, *Paura, legame sociale, ordine politico in Thomas Hobbes*, in G.M. Chiodi, R. Gatti (a cura di), *La filosofia politica di Hobbes*, FrancoAngeli, Milano 2009, pp. 65-79.

¹³ Cfr. A. Silvers, *Feminism and Disability*, in L.M. Alcoff, E.F. Kittay (eds.), *The Blackwell Guide to Feminist Philosophy*, Blackwell, Oxford 2007, pp. 131-142.

¹⁴ Cfr. M.A. Fineman, *The Neutered Mother, the Sexual Family and other Twentieth Century Tragedies*, Routledge, New York 1995.

¹⁵ E.F. Kittay, *La cura dell'amore. Donne, uguaglianza, dipendenza*, cit., p. 52.

¹⁶ Cfr. E.F. Kittay, *At the Margins of Moral Personhood*, "Ethics" 116, 2015, pp. 100-131.

solamente di recente la vulnerabilità è stata “presa sul serio” nell’ambito della riflessione scientifica e della discussione pubblica, soprattutto con riferimento alla sua dimensione ontologica, dunque quale aspetto ineliminabile della soggettività, che può tuttavia trovare declinazioni universalistiche (siamo tutti vulnerabili) e particolaristiche (siamo vulnerabili in modo diverso). Nel dibattito contemporaneo essa rinvia inoltre all’alterità, all’interazione, al riconoscimento reciproco, all’interdipendenza, poiché, in quanto caratteristica degli esseri umani, suggerisce la decostruzione del soggetto, con la sua autonomia e indipendenza, a favore di una nuova concezione fondata sulla relazionalità.¹⁷

Proprio in quanto nozione duttile, da una parte, a dar conto delle diverse precarietà, fragilità, insicurezze, minacce e rischi che contraddistinguono l’epoca contemporanea e, dall’altra, ad affrontare una serie di problemi rilevanti sul versante delle garanzie dei diritti delle persone e del ruolo delle istituzioni, Kittay si impegna a declinare la nozione di vulnerabilità nel caso specifico del lavoro di cura, al di là della tradizionale etica della cura¹⁸. In risposta alla vulnerabilità, il *dependency work* crea dei legami non imposti, ma neppure scelti volontariamente, che sono più diffusi nelle vite umane di quanto si pensi, che determinano l’esistenza di ciascuna persona e che ne definiscono anche lo statuto morale:

quando veniamo meno nel soddisfare un bisogno primario, pur trovandoci nella posizione esclusiva di poter rispondere a quel bisogno, ci chiediamo che tipo di persone siamo.¹⁹

La vulnerabilità umana diventa così la condizione di una reazione che decide della nostra umanità e fonda, senza presupporlo, lo stesso senso morale, svincolato dunque da ragioni deontologiche e da calcoli utilitaristici. Se, da un lato, la vulnerabilità umana diventa emblematica in casi eccezionali e accidentali quali le più gravi forme di disabilità, dall’altro, il fatto di costituire la dimensione ontologica normale e strutturale dell’antropologia, porta Kittay a indicare una dimensione dell’essere umano in cui questi aspetti risultano presenti: quella del figlio. Proprio l’“essere figlio” «dice molto riguardo la relazione forgiata tramite la cura di una persona dipendente vulnerabile e il valore che questa relazione dona a chi è accudito e a chi accudisce. Tale legame è presente ovunque nella società umana ed è fondamentale per la nostra umanità quanto qualsiasi

¹⁷ O. Giolo, B. Pastore, *Vulnerabilità. Analisi multidisciplinare di un concetto*, cit., pp. 11-12.

¹⁸ Cfr. E. Pulcini, *La cura del mondo. Paura e responsabilità nell’età globale*, Bollati Boringhieri, Torino 2009 e J. Braga, M. Santiago de Carvalho (eds.), *Philosophy of Care. New Approaches to Vulnerability, Otherness and Therapy*, Springer, Cham 2021.

¹⁹ E.F. Kittay, *La cura dell’amore. Donne, uguaglianza, dipendenza*, cit., p. 112.

altra caratteristica invocata dai filosofi come ‘distintamente umana’²⁰. Una qualsiasi rivendicazione basata sull’“essere figlio” evita di partire da un presupposto individualistico e pertanto costituisce un’alternativa alle concezioni dominanti delle varie teorie politiche liberali, proprio perché non si appella a una proprietà del singolo, ma a una proprietà che pertiene ad ogni persona in virtù di un’altra persona, cioè di una relazione. Rispetto alle numerose teorie sulla vulnerabilità disponibili, Kittay fa di questa nozione una cifra antropologica e, pur evitando qualsiasi riferimento natura umana, non disdegna di considerarla fonte di una “dignità inalienabile”²¹.

Nel pensiero contemporaneo autori quali Hans Jonas²² e Hanna Arendt²³ hanno già fatto ricorso alla dimensione del “figlio”, ma con l’obiettivo principale di fondare un’etica della cura e di prospettare una filosofia politica basata su di essa. Kittay ne sottolinea piuttosto il valore archetipo e universale: essere figlio per lei significa che la realizzazione avviene solo nella dipendenza, meglio che le potenzialità umane possono compiersi solo in un contesto di ineluttabile vulnerabilità.

3. Definire il *dependency worker*

L’incisività e la profondità con cui Kittay presenta la nozione di vulnerabilità, tanto da farne una cifra ontologica propria dell’essere umano, le consentono di prendere in considerazione la prospettiva di coloro che si prendono cura degli altri e, di conseguenza, le permettono di tematizzarne il ruolo sociale con una forza inedita e da una prospettiva particolare rispetto alle etiche della cura più diffuse. I *dependency workers* sono pertanto i veri protagonisti della riflessione di Kittay ed è proprio a loro che l’autrice presta la sua voce. Una volta assunta la vulnerabilità umana come un mero e “bruto” dato, l’impegno dei *dependency workers* risulta necessario e fondamentale; eppure tale impegno risulta socialmente poco apprezzato, ma soprattutto filosoficamente muto, nel senso che risulta ignorato, meglio neppure visibile, dalle teorie della giustizia prevalenti. Addirittura, al *dependency work* non viene riconosciuto lo statuto di autentico lavoro o, al più, viene relegato nella categoria delle attività non produttive caratteristiche del terzo settore. Questo aspetto non è, secondo Kittay, acciden-

²⁰ Ivi, p. 43.

²¹ Cfr. Ivi, p. 123.

²² Cfr. H. Jonas, *Organismo e libertà. Verso una biologia filosofica*, a cura di P. Becchi, tr. it. di A. Patrucco Becchi, Einaudi, Torino, 1999.

²³ Cfr. H. Arendt, *La vita della mente* [1978], a cura di A. Dal Lago, tr. it. di G. Zanetti, il Mulino, Bologna 1987 e S. Zucal, *Filosofia della nascita*, Morcelliana, Brescia 2017.

tale, bensì è l'esito inevitabile di quell'antropologia riduttiva presupposta da ogni forma di contrattualismo in cui i contraenti sono concepiti quali individui che agiscono per un forte interesse personale, dotati di razionalità e disinteresse reciproco, tutti nella identica posizione di potersi impegnare in relazioni morali reciproche. Kittay²⁴ annovera tra questi percorsi anche quello di Rawls²⁵, pur riconoscendo che il suo “principio di differenza” è un tentativo, per quanto insufficiente, di provare a rispondere alle istanze di chi ha più bisogno, per avere una *chance* in più di includere le esigenze di giustizia di tutti²⁶. L'insufficienza di tale principio sarebbe imputabile all'applicazione troppo limitata che Rawls ne fa nella sua teoria sulla giustizia²⁷. Proprio in quanto (1) la condizione di vulnerabilità non è contemplata nello schema rawlsiano, (2) le esigenze della relazionalità non rientrano nella lista dei beni primari da tutelare, (3) le capacità morali dei contraenti non includono la capacità di rispondere alla vulnerabilità con la cura e (4) nella posizione originaria i *dependency workers* sono esclusi dalla stipula del contratto sociale in quanto la loro libertà e la loro capacità di cooperazione sociale sono estremamente limitate, le conseguenze sono ineludibili:

io affermo che chi si trova all'interno delle relazioni di dipendenza finisce all'esterno del perimetro concettuale dell'egualitarismo di Rawls.²⁸

La critica mossa da Kittay a Rawls è emblematica della sua postura antropologica e dell'alternativa politica alla quale tende:

le politiche sociali portano avanti la finzione del cittadino visto come adulto sano e autonomo, dotato di «pieno funzionamento», come direbbe Rawls, e per il quale la giustizia richiede quella reciprocità possibile con quelli che si trovano nelle stesse condizioni. La dipendenza è concepita soltanto come un'infrazione di questa norma. Le politiche che se ne occupano sono irrilevanti e insufficienti.²⁹

Nella letteratura femminista anglosassone³⁰, dove tale termine è stato coniato, il *dependency work* indica il lavoro di accudire chi è dipen-

²⁴ Cfr. E.F. Kittay, *La cura dell'amore. Donne, uguaglianza, dipendenza*, cit., pp. 135-180.

²⁵ Cfr. J. Rawls, *Una teoria della giustizia* [1971], a cura di S. Maffettone, tr. it. di U. Santini, Feltrinelli, Milano 2017.

²⁶ Cfr. J. Rawls, *La giustizia come equità. Una riformulazione* [2001], a cura di S. Veca, tr. it. di G. Rigamonti, Feltrinelli, Milano 2002.

²⁷ Cfr. J. Rawls, *Liberalismo politico. Nuova edizione ampliata* [1992], tr. it. di A. Ferrara, G. Rigamonti, C. Spinoglio, P. Palminiello, Einaudi, Torino 2012.

²⁸ E.F. Kittay, *La cura dell'amore. Donne, uguaglianza, dipendenza*, cit., p. 143.

²⁹ Ivi, p. 255.

³⁰ Cfr. C. Gilligan, *Con voce di donna. Etica e formazione della personalità* [1982], tr. it. di A. Bottini, Feltrinelli, Milano 1987.

dente e, infatti, Kittay dichiara di utilizzarlo per indicare «il prendersi cura di chi è per forza di cose dipendente»³¹. La studiosa non ha remore a considerarlo un vero e proprio lavoro, emancipandolo fin da subito dall'ambito di un'etica della cura di ambito privatistico. In quanto lavoro vero e proprio, esso pone questioni di etica pubblica e, in particolare, di giustizia sociale. I soggetti che si fanno carico di questo lavoro sono i *dependency workers*, cioè «coloro che si occupano delle persone dipendenti dalle cure di un altro per soddisfarne i bisogni più basilari»³². Poiché, almeno nelle società occidentali contemporanee, tale lavoro continua ad essere prerogativa delle donne, e delle donne appartenenti a classi sociali più svantaggiate, il *dependency work* solleva domande circa la parità di genere, ma soprattutto pone seri interrogativi sulla giustizia, riconducibili alla “critica della dipendenza”, così richiamata da Kittay:

tale critica della dipendenza è una critica femminista dell'uguaglianza, e sostiene che la concezione della società vista come associazione di uguali maschera o occulta ingiuste dipendenze, legate all'infanzia, alla vecchiaia, alla malattia e alla disabilità. Finché siamo dipendenti, non siamo in condizione, in termini di uguaglianza, di entrare in competizione per i beni della cooperazione sociale. E coloro che si prendono cura di chi è dipendente, che devono accantonare i propri interessi per occuparsi di chi è totalmente alla loro mercé, entrano in questa competizione svantaggiati.³³

Il prendersi cura degli altri, per come sono impostate le società liberali occidentali, comporta un serio svantaggio nell'agone pubblico. Nel passo citato Kittay utilizza il termine “dipendenza”, anziché “vulnerabilità”, tradendo così la difficoltà che incontra nel tentare di definire chi sia esattamente un *dependency worker*. In un'accezione troppo inclusiva esso si limita a indicare il lavoro di chi presta una professione funzionale allo svolgimento di un'altra professione, considerata migliore per riconoscimento sociale, retribuzione, titolo di studio e così via. In questo caso ci si limita a sottolineare l'interdipendenza tra tutte le professioni ed eventualmente a criticarne la gerarchizzazione, ma resta lapalissiano che la dipendenza di un capo dalla sua segretaria, dei docenti dagli operatori scolastici, degli ingegneri dagli operai, dei cuochi dai camerieri, dei colletti bianchi dalle donne delle pulizie e così via possano comunque essere fonte di rivendicazioni sociali³⁴. Per delineare il profilo del *dependency worker*, Kittay avverte l'esigenza di una definizione più restrittiva e, per questo, si concentra su quei casi in cui la vulnerabilità è più evidente: il

³¹ E.F. Kittay, *La cura dell'amore. Donne, uguaglianza, dipendenza*, cit., p. XXX.

³² Ivi, p. XVIII.

³³ Ivi, p. XXXII.

³⁴ Cfr. Ivi, p. XXXIV.

dependency work è tale solo nel caso in cui presta aiuto a una persona incapace di provvedere a se stessa. Il punto da sottolineare è che Kittay indugia sulla persona disabile, ma, come abbiamo chiarito nel paragrafo precedente, in realtà ha in mente categorie più estese che comprendono gli anziani, persone affette da malattie invalidanti, anche se solo temporanee, e soprattutto i bambini piccoli. Che i bambini piccoli abbiano necessità di un *dependency worker* per il loro sviluppo, oltre che per la loro sopravvivenza, è proprio ciò che rende, meglio che dovrebbe rendere, la categoria dei *dependency workers* oggetto di particolare attenzione da parte della società. Ogni essere umano ha avuto bisogno di un *dependency worker* per essere colui che è diventato, oltre che a poterne aver bisogno per situazioni contingenti, seppur non marginali: questo elemento costitutivo del tessuto sociale fa sì che ogni persona adulta abbia contratto un vero e proprio debito verso questa categoria, debito che le etiche della cura spesso fanno fatica a cogliere.

Una seconda sfida a cui Kittay deve far fronte nel definire il *dependency worker* è costituita dal fatto che in alcuni casi esso è una vera e propria professione (più o meno) retribuita, in altri casi si tratta di un’assistenza (più o meno volontaria) prestata per ragioni affettive, come nel *mothering*, termine inglese che indica le cure prestate da un genitore (solitamente la madre) nei confronti di un figlio piccolo³⁵ e, per estensione, di un familiare (nella maggior parte dei casi donna) nei confronti di un parente stretto in situazione di non autosufficienza (ad esempio, genitori anziani e conviventi disabili).

La difficoltà a fornire una lista di condizioni necessarie e sufficienti a stabilire in via definitiva chi sia un *dependency worker* è sicuramente una criticità, ma è piuttosto nota nell’ambito delle etiche della cura che, in diversi casi, evitano di presentarsi come autentiche teorie morali, intese come sistemi razionali alternativi ai paradigmi più consolidati³⁶. Il fatto che Kittay non intenda offrire una definizione puntuale del *dependency worker*, anche se alcune indicazioni in tal senso vengono offerte suggerendo che il lavoro propriamente di cura è quello che fa fronte a bisogni “essenziali” delle persone ed è erogato da persone con una scarsa specializzazione professionale, risulta però problematico nella misura in cui l’autrice ambisce a proporre un patto sociale alternativo a quello liberale che sia in grado di riconoscere e dare il giusto valore al *dependency work*. Ciononostante, stabilire criteri certi per distinguere il *dependency work* in senso stretto – con le sue peculiarità di disinteresse e orientamento all’altro – dal *dependency work* in senso largo – che arriva a includere tutte quelle professioni che promuovono o

³⁵ Cfr. F.P. Adorno, *Gli obblighi della cura. Problemi e prospettive delle etiche del care*, cit., pp. 45-50.

³⁶ Cfr. Ivi, p. 65.

incrementano le capacità degli altri – è un'impresa destinata al fallimento, in quanto «il lavoro di provvedere a una persona dipendente non è ben circoscritto»³⁷, proprio in quanto riguarda le caratteristiche legate alla posizione sociale riconosciuta a chi lo svolge.

Il *dependency worker* si definisce meglio attraverso le sue azioni, caratterizzate dall'essere altruistiche, dal portare benefici ad altri a scapito di se stessi, dal non rientrare negli obblighi contrattuali: in una parola dall'essere «azioni supererogatorie»³⁸. Anche in questo caso, non risulta chiaro se tali caratteristiche debbano essere compresenti o meno, poiché oscillano tra l'appartenere a chi svolge un *dependency work* per ragioni professionali e a chi lo svolge per motivi affettivi. Ciò su cui Kittay insiste è che, in entrambi i casi, il lavoro assistenziale, nei fatti, non solo ha uno scarso riconoscimento sociale che lo svaluta, ma crea autentica vulnerabilità in chi lo presta, poiché ne limita le possibilità. Ecco perché un'analisi filosofica del lavoro assistenziale centrato sul soggetto che lo presta oltre a chi se ne avvale, deve utilizzare apparati di analisi concettuale complementari rispetto a quelli offerti delle filosofie della cura.

4. Caratteristiche del *dependency worker*

La discussione del paragrafo precedente ha mostrato come le istanze del *dependency work* siano generalmente considerate una questione privata e siano fortemente connotato dal genere, due fattori che hanno contribuito a eludere qualsiasi efficace riflessione intorno ad una equa distribuzione del *dependency work* sia tra generi, sia tra classi sociali. L'urgenza, per Kittay, diventa pertanto quella di spostare i problemi legati a questo tipo di lavoro dal piano della morale privata a quello dell'etica pubblica, con la speranza di apportare dei cambiamenti anche nelle politiche sociali. In altri termini, il *dependency work*, oltre ad essere un oggetto di studio dell'etica della cura³⁹, dovrebbe avere un ruolo centrale in ogni teoria di giustizia sociale. Per comprendere questi passaggi auspicati da Kittay è utile individuare alcune caratteristiche del *dependency worker*.

4.1 L'«io trasparente»

La prima, concerne l'identità personale, definita con un neologismo che sposta l'attenzione da una identità centrata sulla volontà di chi presta

³⁷ E.F. Kittay, *La cura dell'amore. Donne, uguaglianza, dipendenza*, cit., p. 71.

³⁸ Ivi, p. 95.

³⁹ Cfr. A. Loretoni, *Etica della cura e vulnerabilità degli individui*, «La società degli individui», 63, 2018, pp. 77-90.

la cura a una identità che paga lo scotto del legame a stretto filo con colui verso cui la cura è prestata: l’“io trasparente”. Si tratta dell’io di chi, per periodi più o meno lunghi e per ragioni affettive o professionali, si prende cura di un’altra persona durante la sua crescita o che ne ha reso concretamente possibile la realizzazione, la *floritura*, per usare un termine caro all’etica della virtù⁴⁰, nei momenti in cui non era capace di badare a se stesso. La dizione “io trasparente” è coniata da Kittay in polemica con l’io cartesiano e, più in generale, con l’individualismo liberale, ma anche con i dispersivi tentativi di circoscrivere un “io femminile”⁴¹. Esso esprime in maniera eloquente la condizione umana e, in particolare, l’atteggiamento di chi, guardando se stesso e volendo realizzare la propria persona, vede prima di tutto l’altro, i suoi bisogni, le sue esigenze, le sue aspirazioni e le antepone, volente o nolente, alle proprie⁴².

Nel concreto, l’“io trasparente” si incarna prevalentemente nell’“io” delle donne: sono, infatti, in prevalenza i soggetti femminili a occuparsi della vulnerabilità umana: sono le donne a prendersi cura dei figli, dei genitori, dei mariti, degli anziani di casa e dei familiari disabili e sono soprattutto le donne a farlo, oltre che per professione, per ragioni affettive, cioè politicamente invisibili. La trasparenza dell’io offre così lo spunto per una riflessione sul femminismo e l’occasione per ripensarne i presupposti. In particolare, Kittay, pur riconoscendo i meriti di quel femminismo che ha fatto della nozione di uguaglianza la sua bandiera, ritiene che sia ormai il momento di riconoscere gli ideali liberali di imparzialità, neutralità, uguaglianza insufficienti a promuovere una visione egualitaria tra tutti gli esseri umani⁴³.

Il nesso tra trasparenza dell’io e questione femminile aggiunge una accezione politica all’“io trasparente”, attribuendogli un ulteriore significato. L’“io” del *dependency worker* è “trasparente” perché non attira alcuno sguardo, oltre ad emettere una voce non udibile⁴⁴. In una società in cui l’individualismo astratto del liberalismo è incapace di vedere le istanze reali della condizione umana, cioè della sua vulnerabilità, né sa comprendere il valore delle dipendenze e delle reciprocità asimmetriche che si instaurano durante l’intero arco dell’esistenza umana, non può stupire che l’io del *dependency worker* sia ignorato: è arduo rendersi conto a sufficienza delle sue aspirazioni latenti, delle sue frustrazioni, dei meccanismi psicologici che lo investono, della

⁴⁰ Cfr. A. Campodonico, M. Croce, M.S. Vaccarezza, *Etica delle virtù. Un’introduzione*, Carocci, Roma 2017.

⁴¹ Cfr. E.F. Kittay, *La cura dell’amore. Donne, uguaglianza, dipendenza*, cit., pp. 92-93.

⁴² Cfr. Ivi, p. 95.

⁴³ Cfr. Ivi, pp. 4-9.

⁴⁴ Cfr. F.P. Adorno, *Gli obblighi della cura. Problemi e prospettive delle etiche del care*, cit., pp. 20-22.

sua stanchezza, della sua solitudine e, in ultima analisi, dei suoi diritti. Troppo sbrigativamente tali aspetti vengono ricondotti a qualche processo di *burnout*, se il lavoro di dipendenza è una professione, o a effetti collaterali più o meno voluti o inattesi, ma comunque frutto di una scelta volontaria e deliberata, se il lavoro di dipendenza viene svolto per ragioni affettive. L'analisi dell'identità del *dependency worker* proposta da Kittay si discosta da quella offerta dalle etiche della cura che prestano attenzione alle peculiarità di questo particolare tipo di lavoratore promuovendo un dibattito che, oscillando tra empatia e simpatia⁴⁵, si sofferma sulla pazienza, sull'umiltà, sul coraggio, sulla tenacia, sulla fiducia e sul rispetto⁴⁶, ma tralascia le conseguenze che porta con sé, rese ancora più drammatiche dal disinteresse da parte degli agenti politici.

4.2 Peculiarità paradigmatiche

I requisiti pratici indispensabili e le esigenze morali essenziali di un *dependency work* efficace, adeguato e correttamente impostato consistono nell'equilibrio di tre aspetti, già raccolti da Jane Martin⁴⁷ nell'acronimo CCC: *care*, *concern* e *connection*. Kittay li declina come segue:

è il lavoro di occuparsi degli altri nella loro condizione di vulnerabilità – cura. Mantiene legami tra intimi o crea esso stesso intimità e fiducia – relazione. Così che i legami affettivi – preoccupazione – in genere sostengano tale relazione, anche quando il lavoro implica uno scambio economico.⁴⁸

Nell'accezione forte con cui Kittay declina il *dependency work*, la dimensione di reciprocità viene esclusa, proprio perché l'assistenza è nei confronti di persone incapaci a badare a se stesse, il che elimina la possibilità di uno scambio simmetrico, se non in una dimensione diacronica, in cui talvolta i ruoli addirittura si invertono, finendo comunque con il confermare l'asimmetria nel momento specifico della prestazione. L'assenza di reciprocità concerne la gratuità dell'assistenza e la mancanza di calcolo, ma comprende la possibilità di una crescita personale, nella misura in cui colui che presta assistenza affina il suo senso morale. Inoltre, la tendenza a professionalizzare il *dependency work* se, da un lato, ha avuto indubbi benefici, dall'altro, tende a renderlo poco riconoscibile

⁴⁵ Cfr. Ivi, pp. 96-100.

⁴⁶ Cfr. Ivi, pp. 44-45.

⁴⁷ Cfr. J. Martin, *Transforming Moral Education*, in M.M. Brabeck (ed.), *Who Cares? Theory, Research, and Educational Implications of the Ethic of Care*, New York, Praeger, 1989, pp. 183-196.

⁴⁸ E.F. Kittay, *La cura dell'amore. Donne, uguaglianza, dipendenza*, cit., p. 56.

in quanto tale: una volta diventato una professione, le logiche di questa dimensione prevalgono e, per converso, là dove resta mero *dependency work* la sua rilevanza pubblica diminuisce fino a scomparire e l’aspetto privato lo occulta. Infine, un’ultima caratteristica riguarda la tendenza del *dependency work* ad essere appannaggio di una classe lavoratrice a sé. Il fatto che esso risulti prevalente in una certa classe sociale definita attraverso età, reddito, razza, etnia e genere è un dato su cui intervenire, una caratteristica attuale del *dependency work* che, a differenza delle precedenti, può e deve essere rivista. Purtroppo, chiosa Kittay, «la visione illuministica lascia incontrastato il ruolo delle donne come *dependency workers*»⁴⁹. A partire dal *dependency work*, la questione della donna che se ne incarica fa problema a prescindere – in particolare, a prescindere dalla scelta (presunta) libera con cui si dedica a tale lavoro – in quanto prelude comunque a una forma di ingiustizia sociale. Al contrario, le riflessioni sul “lavoro di cura” guardano con sospetto solo all’ambito delle richieste superrogatorie dovute a pressioni sociali⁵⁰ tanto che, almeno in alcune forme, hanno la tendenza a valorizzare la specificità del femminile⁵¹ finendo con eludere le istanze emancipatorie delle critiche femministe.

4.3 Rischi patologici

All’ingiustizia sociale che accompagna il *dependency worker* si aggiungono dei rischi patologici che accompagnano la sua attività. Tali rischi sono ben conosciuti dai teorici del *care* che li declinano come eccessi di paternalismo a cui ogni lavoro di cura può andare incontro⁵² e che affliggono il soggetto della cura. Kittay, in maniera speculare, considera i rischi propri del *dependency worker*, cioè quelli che restano, per così dire, a suo carico. Peculiare del *dependency work* è la sfera dell’emotività, su cui Kittay spende parole incisive:

la delicatezza dell’equilibrio e del controllo emotivo richiesta da ambo le parti e l’immenso investimento di energie emozionali necessarie per realizzare il *dependency work* sono segno di una relazione profondamente

⁴⁹ Ivi, p. 74. La critica al paradigma illuminista si ritrova espressa in termini analoghi anche in E. Pulcini, S. Bourgault, *Introduzione a una mappa affettiva della relazione di cura*, in Idd. (a cura di), *Cura ed emozioni. Un’alleanza complessa*, il Mulino, Bologna 2018, pp. 7-22.

⁵⁰ Cfr. F.P. Adorno, *Gli obblighi della cura. Problemi e prospettive delle etiche del care*, cit., pp. 51-42.

⁵¹ Cfr. Ivi, pp. 50-52.

⁵² Cfr. Ivi, p. 54.

affettiva che esige fiducia. Proprio per la rilevanza sia dell'affetto, sia della fiducia entro queste relazioni, i legami che ne derivano sono tra i più importanti di cui facciamo esperienza. Spesso sembra che infondere la cura dell'amore all'interno di simili relazioni [...] allarghi i confini di noi stessi, concedendo così spazio a un legame emotivo che risulta particolarmente potente.⁵³

Ciò che viene dettagliato con un linguaggio così rispettoso e in una prospettiva dai tratti edificanti, rischia però di trasformarsi in qualcosa di dannoso e di patologico. Ogni *dependency work* comporta una disuguaglianza di potere tra i soggetti coinvolti che potenzialmente può trasformarsi in un esercizio di dominio, cioè in un esercizio illegittimo di potere. La corruzione del rapporto di potere può dipendere tanto dal soggetto che presta la cura, quanto dal soggetto che ne fruisce. Chi si occupa della vulnerabilità dell'altro si trova costantemente in una posizione da cui può abusare del suo ruolo, talvolta anche senza volerlo. Ma anche chi dipende può esercitare una certa tirannia millantando falsi bisogni, generando sensi di colpa, creando frustrazioni eccessive, favorendo ricatti psicologici, sfruttando l'empatia altrui per tornaconti egoistici. Anche le passioni e le emozioni legate alla compassione possono avere risvolti negativi⁵⁴. Alle relazioni che si instaurano in un *dependency work* pertengono delle conflittualità latenti, quando non manifeste:

L'interesse del *dependency worker* rimane legato a quello di chi è preso in carico, anche se si pone in contrasto con esso. La sua libertà di crearsi i propri obiettivi e dare espressione ai propri desideri è limitata non solo dagli obiettivi e dai desideri di altri esseri autonomi come lui; ma anche, e forse in modo più radicale, dalle esigenze e dalle aspirazioni di chi dipende da lui.⁵⁵

Affinché queste relazioni già di per sé asimmetriche non diventino conflittuali e non finiscano per trasformarsi in patologiche, ma si conservino sane, efficienti e generative è necessario che la società garantisca l'accesso a quelle risorse necessarie per il mantenimento e il benessere sia della persona non autosufficiente, sia della persona che se ne fa carico. Ancora una volta emerge la necessità di affrontare gli obblighi morali di cui è necessario farsi carico nei confronti dei *dependency workers*.

⁵³ E.F. Kittay, *La cura dell'amore. Donne, uguaglianza, dipendenza*, cit., p. 64.

⁵⁴ Cfr. E. Pulcini, *Quali emozioni motivano alla cura?*, in E. Pulcini, S. Bourgault (a cura di), *Cura ed emozioni. Un'alleanza complessa*, cit., pp. 25-47. (Pulcini 2018).

⁵⁵ E.F. Kittay, *La cura dell'amore. Donne, uguaglianza, dipendenza*, cit., p. 75.

5. Dal *welfare* alla *doulia*

Circoscritta la natura del *dependency worker* e assodata la necessità della sua funzione nel consesso umano in forza della vulnerabilità tipica dei suoi membri, è necessario considerare il ruolo che la società dovrebbe riconoscergli e che invece ha per troppo tempo misconosciuto, se non ignorato anche negli autori più sensibili a questo aspetto, quali John Stuart Mill⁵⁶. Questo passaggio viene così descritto:

l'indagine comincia da un'auto-comprensione delle nazioni democratiche e liberali concepite come associazioni di individui eguali, liberi e indipendenti. Voglio sfidare questa auto-comprensione poiché siamo tutti, in diversi momenti, dipendenti. Molti di noi, soprattutto le donne, devono anche occuparsi delle esigenze di chi dipende. Il concetto che tutti noi fungiamo, almeno idealmente, da cittadini liberi e uguali, non è smentito solo dalla realtà empirica: anche teoricamente non è un concetto abbastanza ampio da comprendere tutti.⁵⁷

Con queste parole, Kittay riformula la classica critica femminista della dipendenza nei termini di una “critica della dipendenza all’uguaglianza”⁵⁸. Una richiesta di uguaglianza che si accontenta di rivendicare i beni appartenenti a un gruppo dominante senza prendere in considerazione il *dependency work* è destinata inevitabilmente e tragicamente a scaricare il peso di persone che sono entrate a far parte del gruppo dominante su altri soggetti che hanno meno potere: ciò si sta verificando con l’attribuzione del *dependency work* a giovani donne immigrate dai paesi via via più poveri⁵⁹. Soltanto quando l’uguaglianza saprà rispondere oltre che a esigenze di giustizia e a esigenze proprie della cura e della relazione, anche alle esigenze tipiche della vulnerabilità, allora l’uguaglianza conquistata da alcuni non andrà a scapito di altri, ma distribuirà gli oneri della vulnerabilità umana in modo equo su tutta la popolazione. Kittay indica il paradigma morale più adatto:

⁵⁶ Cfr. Ivi, p. 16.

⁵⁷ Ivi, p. 10.

⁵⁸ Cfr. Ivi, pp. 28-34.

⁵⁹ Circa questo fenomeno si veda, oltre a E.F. Kittay, *The Moral Harm of Migrant Carework*, “Philosophical Topics”, 37, 2009, pp. 53-73, anche J. Andall, *Gender, Migration and Domestic Service. The Politics of Black Women in Italy*, Aldershot, Ashgate 2000; Id. (ed.), *Gender and Ethnicity in Contemporary Europe*, Berg, Oxford and New York 2003; K. Calavita, *Immigrants at the Margins. Law, Race and Exclusion in Southern Europe*, Cambridge University Press, Cambridge 2005; M. Cvajner, *Sociologia delle migrazioni femminili*, il Mulino, Bologna 2018; Id., *Soviet Signoras. Personal and Collective Transformations in Eastern European Migration*, University of Chicago Press, Chicago 2019 e J. Henshall Momsen (ed.), *Gender, Migration and Domestic Service*, Routledge, London and New York 1999.

la caratteristica dell'io morale, l'asimmetria della relazione, la posizione non imparziale dei suoi componenti, la sua natura non volontaria rendono le esigenze morali della relazione di dipendenza più soggette a un'etica della cura che a una moralità basata sui diritti o sull'utilitarismo.⁶⁰

L'etica della cura ha un doppio vantaggio rispetto alle morali deontologiche, a quelle utilitariste⁶¹ e a quelle basate sull'appello ai diritti umani⁶²: da un lato, considera il soggetto morale intrinsecamente relazionale; dall'altro, privilegia un ragionamento morale contestuale e responsabile, lontano da un calcolo di vantaggi e diritti. Ma Kittay non la ritiene ancora sufficiente.

Le concezioni classiche dell'uguaglianza, pur nella loro diversità, sono accomunate dal rivendicare un qualche trattamento, delle opportunità, un certo benessere, delle risorse, dei beni sociali o delle capacità in forza di una proprietà che ci caratterizza come individui, sia essa la razionalità, determinate passioni oppure il possesso di certi organi. Questo tipo di uguaglianza viene perseguita, per altro con dubbi risultati, attraverso il ricorso al *welfare* nella sua triplice accezione di "parità di risorse", "pari opportunità al *welfare*" o "parità di *welfare*"⁶³, cioè di assegnazione di qualcosa a un individuo grazie a qualche suo merito intrinseco. Esaminando le politiche americane del *welfare* perseguite in modo diverso da conservatori e da progressisti, Kittay⁶⁴ nota che, in ogni caso, le giustificazioni si rifanno alle medesime strutture filosofiche, in base alle quali la persona assistita dal *welfare* è vista come qualcuno che può godere della piena cittadinanza soltanto ricoprendo il ruolo di percettore di reddito indipendente, misconoscendo così alla radice le caratteristiche dei *dependency workers* e le loro esigenze. Kittay invita pertanto a un cambio radicale di paradigma che non si limita ad adottare la prospettiva di un'etica della cura, ma fa appello a una diversa antropologia, in coerenza con quanto visto nei paragrafi precedenti. Le persone devono essere considerate non in quanto individui, cioè sulla base della loro razionalità e dei loro interessi, bensì per il loro essere coinvolte in relazioni di accudimento. Per continuare a parlare di uguaglianza, dobbiamo concedere che essa vada valutata a partire dalle relazioni in cui le persone sono coinvolte e non dalle preferenze degli individui. In questo nuovo paradigma, ispirato

⁶⁰ E.F. Kittay, *La cura dell'amore. Donne, uguaglianza, dipendenza*, cit., p. 97).

⁶¹ Cfr. Id., *The Ethics of Philosophizing: Ideal Theory and the Exclusion of People with Severe Cognitive Disabilities*, in L. Tessman (ed.), *Feminist Ethics and Social and Political Philosophy: Theorizing the Non-Ideal*, Springer, Cham 2009, pp. 121-146.

⁶² Cfr. J. Morris, *Impairment and Disability: Constructing an Ethics of Care That Promotes Human Rights*, "Hypatia", 16, 2001 pp. 1-16.

⁶³ Cfr. E.F. Kittay, *La cura dell'amore. Donne, uguaglianza, dipendenza*, cit., p. 44.

⁶⁴ Cfr. Ivi, pp. 215-245.

e, in certa misura favorito dalle teorie del *care*, ma non necessariamente richiesto da esse, cambia la richiesta stessa di uguaglianza

la domanda dell’uguaglianza basata sulla relazione non è: quali diritti mi spettano in virtù del mio stato di ‘eguale’, diritti tali da essere conformi a quelli di tutti gli altri individui che godono del mio stesso stato di ‘egualità’? Invece la domanda è: quali sono le mie responsabilità verso coloro con cui mi trovo in relazioni particolari e quali sono le loro responsabilità nei miei confronti, cosicché io possa essere assistito e vedere accolte le mie esigenze anche quando io mi occupo e rispondo alle esigenze di chi dipende da me?⁶⁵

Come modello – alternativo al *welfare* – capace di comprendere le esigenze della vulnerabilità e di venire incontro alle esigenze delle necessarie interrelazioni umane in una prassi sociale e politica adeguata, Kittay propone quello della *doula* greca⁶⁶. Le ragioni di questa scelta non sono evidenti, considerando che la *doula* indicava in origine una condizione servile e di schiavitù⁶⁷ che contrasta con la proposta di Kittay. L’autrice è consapevole della sua provocazione, ma ritiene che il riferimento alla *doula* sia funzionale ad evitare una troppo frettolosa associazione della sua posizione con le etiche della cura⁶⁸. La schiavitù resta una forma di servizio moralmente intollerabile che umilia e disumanizza indipendentemente dalle condizioni in cui è prestata, ma può venire ricondotta, in qualche modo, a un lavoro di cura. Il *dependency work* rappresenta, al contrario, la forma più basilare delle relazioni umane e si insinua nel cuore dell’umanità, ma solo a patto che rispetti i principi, non solo formali, di giustizia e di uguaglianza. Le ragioni per cui viene richiamata la figura della *doula* non fanno riferimento alla condizione servile e alla mancanza di libertà, ma neppure al tipo di attività svolta. Piuttosto Kittay sembra interessata alla *doula* in quanto aveva una funzione sociale stabilita e un

⁶⁵ Ivi, p. 49.

⁶⁶ Questo termine ha avuto diffusione grazie al testo dell’antropologa americana D. Raphael, *The Tender Gift: Breastfeeding: Mothering the Mother – The Way to Successful Breastfeeding*, Prentice Hall, Englewood Cliffs, N.J. 1973 per indicare «one or more individuals, often female, who give psychological encouragement and physical assistance to the newly delivered mother» (p. 172).

⁶⁷ Per una sintetica bibliografia sull’istituto della *doulia* sotto il profilo storico rimando a L. Sichirolo (a cura di), *Schiavitù antica e moderna: problema, storia, istituzioni*, Guida, Napoli 1979; M.M. Mactoux, *Douleia: Esclavage et pratiques discursives dans l’Athènes classique*, Belles Lettres, Paris 1980; I. Biežuńska-Małowist, *La schiavitù nel mondo antico*, Edizioni Scientifiche Italiane, Napoli 1991; M. Falappone, *La schiavitù nel mondo antico*, Laterza, Roma-Bari 1988; P. Hunt, *Slaves, Warfare, and Ideology in the Greek Historians*, Cambridge University Press, New York 2002 e P. Vidal-Naquet, *Il cacciatore nero. Forme di pensiero e forme d’articolazione sociale nel mondo greco antico* [1981], tr. it. di F. Sircana, Feltrinelli, Milano 2006.

⁶⁸ Cfr. E.F. Kittay, *La cura dell’amore. Donne, uguaglianza, dipendenza*, cit., pp. 197-198.

riconoscimento formale tali da garantire una forma di reciprocità transittiva, trasferibile e intergenerazionale. Questi sembrano essere i motivi prevalenti che spingono Kittay ad auspicare la sostituzione del *welfare* nelle società contemporanee con la figura della *doulia*:

proprio come abbiamo avuto bisogno di assistenza per sopravvivere e crescere bene, allo stesso modo è necessario che offriamo ad altri – compresi quelli che prestano le loro cure per lavoro – le condizioni che permettano loro di ricevere l'assistenza di cui hanno bisogno per la sopravvivenza e la prosperità.⁶⁹

La scelta della *doulia* quale sostituto al *welfare* soggiace alle pagine più utopiche di Kittay. Il riferimento al mondo greco resta su un piano piuttosto evocativo e l'argomentazione proposta si fa più convincente solo attraverso un ampliamento della nozione di servizio. L'analogia tra il servizio reso dalla *doula* alla madre che ha partorito, in modo che quest'ultima possa occuparsi direttamente del neonato, e il tipo di servizio reso dai *dependency workers* alla società, risulta comprensibile solo spostandosi dalla circostanza privata dell'assistenza post-parto a una concezione pubblica di cura capace di riconoscere cerchi di reciprocità che si muovono verso l'esterno in direzione di strutture sociali più ampie, dove l'uguaglianza e la giustizia vengono sempre e comunque garantite⁷⁰. In termini concreti:

io propongo che il 'lavoro di dipendenza' venga integrato in un sistema sociale e che si stabilisca una retribuzione universale.⁷¹

6. Un *case study* e prospettive future

Molte riflessioni contenute ne *La cura dell'amore* hanno un debito verso il contesto storico, geografico, culturale e politico nel quale il libro è stato scritto. Volendo privilegiare quegli aspetti filosofici che lo rendono fruibile anche in scenari lontani da quello che lo hanno generato, ho trascurato i dati contestuali più contingenti⁷². Lungi dall'indugiare su una prospettiva utopica dove le suggestioni antropologiche faticano a diven-

⁶⁹ Ivi, p. 195.

⁷⁰ Cfr. Ivi, p. 244.

⁷¹ Ivi, p. 259.

⁷² Per questo aspetto rimando all'esame critico del *Family and Medical Act* contenuto in Ivi, pp. 245-254 e, più in generale, a N. Fraser, L. Gordon, *The Genealogy of Dependency. Tracing a Keyword of the US Welfare State*, in E.F. Kittay, E.K. Feder (eds.), *The Subject of Care. Feminist Perspectives on Dependency*, Rowman and Littlefield Publishers, Oxford 2002, pp. 143-149.

tare prassi vissuta, Kittay è la prima a mettere in guardia il lettore della necessità di correggere il proprio modo di relazionarsi con gli altri, prima di voler cambiare l’ordine del mondo. Ogni cambiamento sociale e politico, per essere incisivo e duraturo, deve avere una solida base antropologica in grado di offrire dei criteri adeguati di giustizia. In Italia la veridicità di tale priorità è stata confermata nel caso delle persone con disabilità: la difficile storia degli “handicappati”⁷³ ha conosciuto una svolta grazie alla legge 104 del 1992⁷⁴, ma solo negli ultimi decenni l’attenzione giurisprudenziale alle persone con disabilità ha messo l’accento sulle “persone”, con solidi argomenti di antropologia filosofica, desunti proprio da Kittay⁷⁵. Limitatamente all’auspicio del “prendersi cura di chi si prende cura”, merita di essere ricordata una proposta di disegno di legge sui “caregiver familiari” avanzata in Senato nel corso della XVII legislatura e discussa ampiamente nel corso della XVIII. Questa proposta ha avuto una certa risonanza poiché ha intercettato sia le vulnerabilità emerse nel corso della prima ondata pandemica di Covid-19⁷⁶, sia quelle tipiche di una popolazione sempre più anziana⁷⁷ segnata da un forte calo demografico⁷⁸. Essa, inoltre, si propone come tentativo, da un lato, di organizzare, centralizzare e dare maggiore sostanza e stabilità a diverse iniziative nate con lo stesso spirito a livello regionale⁷⁹ e, dall’altro, di favorire la ricezione in Italia della *Carta europea della disabilità*⁸⁰. L’introduzione al

⁷³ Cfr. A. Canevaro, A. Goussot (a cura di), *La difficile storia degli handicappati*, Carocci, Roma 2002.

⁷⁴ Il testo completo è scaricabile dal sito < <https://www.gazzettaufficiale.it/eli/id/1992/02/17/092G0108/sg> > cons. il 25 febbraio 2024.

⁷⁵ Cfr. M. Zanichelli, *Persone prima che disabili. Una riflessione sull’handicap tra giustizia ed etica*, Queriniana Brescia 2012 e Id., *Paradigmi di giustizia alla prova della disabilità*, “Rivista di Filosofia del Diritto”, 2, 2013, pp. 205-226.

⁷⁶ Cfr. A. Loretoni, *La cura del mondo comune. Vulnerabilità di individui e istituzioni nella fase della pandemia*, “La società degli individui”, 69, 2020 pp. 136-145 e B. Henry, *Alcune riflessioni sugli scenari della pandemia. Libertà, sicurezza, vulnerabilità*, “La società degli individui”, 69, 2020 pp. 117-126.

⁷⁷ Cfr. S. Amato, *Animali razionali dipendenti? La Carta dei diritti degli anziani e dei doveri della società*, “Politica.eu”, 7, 2021, pp. 15-33.

⁷⁸ Cfr. A. Minello, *Non è un paese per madri*, Laterza, Roma-Bari 2022.

⁷⁹ Si pensi, a mero titolo esemplificativo e limitatamente alla regione Liguria, al *Fondo regionale per il sostegno del ruolo di cura e assistenza del caregiver familiare* stanziato a seguito del Decreto del Dirigente del settore Politiche Sociali n. 8131 del 30 dicembre 2021 e al *Corso di formazione per i caregiver dei pazienti con SLA* promosso dall’ASL3 – Liguria. La più recente iniziativa analoga è la legge *Norme per il riconoscimento ed il sostegno del caregiver familiare* promulgata dalla Giunta Regionale del Lazio l’8 febbraio 2024 che si distingue per venire incontro alle esigenze specifiche dei caregiver più giovani nel delicato periodo della loro formazione professionale, a seguito di una stima per la quale il 7% dei giovani compresi tra i 16 e i 26 anni potrebbe essere un caregiver.

⁸⁰ I criteri per il rilascio della *Carta europea della disabilità* in Italia sono state previste dal DPCM del 6 novembre 2020 e pubblicate dal Consiglio dei Ministri nella Gazzetta

testo consiste in una presa di consapevolezza di quella figura che finora abbiamo indicato, seguendo Kittay, con il termine di *dependency worker*:

assistere una persona cara non autosufficiente ed esserle di aiuto nelle difficoltà di gestione della vita quotidiana costituisce una funzione cardine delle relazioni di convivenza, basate sulla libera scelta e alimentate da motivazioni affettive e sentimentali. Ci sono situazioni in cui questa funzione di aiuto assume connotati di impegno tali da rendere necessari ed opportuni interventi di sostegno da parte delle istituzioni pubbliche, in favore delle persone che si trovano nella condizione di assistere una persona cui sono legate per motivi affettivi o di parentela, quale che sia la loro età, perché affette da patologie invalidanti, anche croniche o degenerative.⁸¹

Non è questa la sede per discutere i dodici articoli di cui si compone questa proposta e le loro modifiche nel corso dei lavori in commissione, dalle quali si evincono le difficoltà economiche e le criticità sociali ad essa connesse e dalla non facile soluzione. Resta l'esigenza – e non è poco – di normare il riconoscimento e la tutela del lavoro svolto dai *caregivers* familiari «che rappresenta un valore sociale ed economico per il Paese»⁸², valore «connesso ai rilevanti vantaggi che dalla sua [del *caregiver*] opera trae l'intera collettività»⁸³.

Il richiamo a questo *case study*, mette in luce il guadagno della proposta di Kittay rispetto alle etiche della cura. La figura del *dependency worker*, pur nelle ambiguità segnalate, è comunque meglio circoscrivibile rispetto alla nozione di cura⁸⁴ e ha il vantaggio di identificare un soggetto bisognoso di tutele giuridiche e garanzie economiche. Questo passaggio dalla sfera etica alla dimensione politica supera la tradizionale contrapposizione tra *care* e giustizia, ha però un alto costo teorico, infatti necessita di accettare lo slittamento dal concetto di dipendenza a quello di vulnerabilità e, soprattutto, di condividere la tesi per cui la

Ufficiale n. 304 del 23 dicembre 2021.

⁸¹ Senato della Repubblica. XVIII Legislatura (2021), *Fascicolo Iter DDL S. 1461 Disposizioni per il riconoscimento ed il sostegno del caregiver familiare*, < https://www.senato.it/leg/18/BGT/Schede/Ddliter/testi/52186_testi.htm > cons. il 25 febbraio 2024, p. 5.

⁸² Ivi, p. 6.

⁸³ Ivi, p. 7. Per gli ultimi sviluppi della discussione in merito rimando alla registrazione del convegno “Nessuno è escluso. Disabilità, una questione di famiglia” tenutosi l'8 novembre 2023 nella Sala Matteotti presso la Camera dei Deputati reperibile all'indirizzo <https://webtv.camera.it/evento/23657>. Per una sensibilizzazione dell'opinione pubblica nei confronti del caregiver segnalò il film *Il mio compleanno* presentato al Festival del Cinema di Venezia nel novembre 2024.

⁸⁴ Cfr. F.P. Adorno, *Gli obblighi della cura. Problemi e prospettive delle etiche del care*, cit., pp. 60-61.

vulnerabilità deve essere gestita, ma non può essere eliminata, neppure in linea di principio. Le motivazioni etiche portate da Kittay difficilmente esauriscono la necessità di giustificare il passaggio dal suo modello antropologico di matrice fenomenologico-descrittiva alla richiesta di un sistema prescrittivo e normativo. Il riferimento alla dimensione del figlio non è così facilmente estensibile anche ad altre forme di vulnerabilità e l’analogia tra il *dependency worker* la figura della *doula* non risulta ancora del tutto convincente. Se occuparsi della vulnerabilità di un neonato obbliga in qualche modo la società, in quanto apertura al futuro, occuparsi di altre forme di dipendenza richiede un altro tipo di giustificazioni che pertengono al riconoscimento e alla dignità umana. Qui la proposta della Kittay andrebbe forse compresa non tanto in dialettica con le etiche della cura, quanto con le etiche delle virtù, aprendo un ulteriore scenario che esula dagli scopi di questo contributo.

Maria Teresa Pacilè*

**Per una rinascita spirituale dell'uomo.
Karl Jaspers, la svolta assiale e il progetto
della *Weltphilosophie***

Abstract

For a spiritual rebirth of man.

Karl Jaspers, the Axial Age and the project of the *Weltphilosophie*

This paper aims to explore Jaspersian notion of “axial revolution”, interpreted as a fundamental turning point in universal human history. In the light of the current debate on the Axial Age and the critical interpretations offered by the comparative sociology (Eisenstadt) and the contemporary historiography (Assman and Bellah), it may be interesting to read the Jaspers' thesis of axiality as constant *spiritual rebirth* within the project of a history of world philosophy (*Weltphilosophie*).

Keywords

Re-birth; Axial Age; Western Thought; Eastern Wisdom; World History of Philosophy

Il nostro tempo vive una profonda instabilità, esistenziale, etica e politica. Vengono alla mente alcuni passaggi di *Die geistige Situation der Zeit*¹, in cui Karl Jaspers affermava: “Ci furono epoche nelle quali l'uomo sentiva il proprio mondo come *permanente* [...]. In confronto a siffatte epoche l'uomo d'oggi [...] è separato dalla propria radice”². All'interno di un tempo di inaspettati cambiamenti, con il conflitto russo-ucraino ai confini dell'Europa e la sempre più violenta guerra israelo-palestinese, sembra che la *crisi* sia l'unico *senso* che la storia possa assumere. Per tentare di dare risposte costruttive a tale incertezza diventa urgente una nuova interrogazione sul passato, fondamentale, come insegna Jaspers, per “accrescere la nostra coscienza del presente”³.

* Università degli Studi di Messina

¹ Cfr. K. Jaspers, *Die geistige Situation der Zeit*, De Gruyter, Berlin 1931; tr. it. *La situazione spirituale del tempo*, Jouvence, Roma 1982.

² Ivi, pp. 30-31.

³ K. Jaspers, *Vom Ursprung und Ziel der Geschichte*, Artemis, Zurigo 1949; tr. it. *Origine e senso della storia*, Comunità, Milano 2014, p. 9.



Vivendo a pieno la *crisi* della Germania del secondo dopoguerra, richiamato a insegnare all'università di Heidelberg dopo essere stato costretto al silenzio sotto la dittatura nazista, Karl Jaspers investe le proprie energie nel tentativo di delineare un progetto condiviso di pace. Quest'ultimo non può essere disgiunto dall'analisi storico-politica della Germania post-bellica, dal riconoscimento delle sue colpe e delle sue responsabilità, necessario per avviare un percorso di redenzione orientato a una *rinascita spirituale*¹. Proprio in tale clima esistenziale emerge l'esigenza di una nuova "coscienza" storica², che muove Jaspers alla scrittura di *Vom Ursprung und Ziel der Geschichte* (1949), in cui scrive:

Vogliamo comprendere la storia come tutto per comprendere noi stessi. La storia è per noi il ricordo, di cui non solo abbiamo conoscenza, ma sulla cui base viviamo. Essa costituisce le fondamenta che sono già poste e a cui rimaniamo legati se vogliamo non disperderci nel nulla, ma prender parte all'essere-umano. Una visione della storia crea lo spazio da cui si desta la nostra coscienza dell'essere umano.³

Questa "rinascita spirituale"⁴, questo desiderio di un "nuovo inizio" intrecciano il discorso jaspersiano a quello di Hannah Arendt ponendo i presupposti teorici, etici e politici della riflessione filosofica sulla natalità⁵. Nel dialogo costante tra il Maestro e l'allieva – documentato in particolar modo dal *Briefswechsel*⁶ – è possibile cogliere un intreccio di

¹ Cfr. K. Jaspers, *Die Schuldfrage*, Schneider, Heidelberg 1946; tr. it. *La questione della colpa. Sulla responsabilità politica della Germania*, Raffaello Cortina, Milano 1996.

² Interessante notare che nello stesso anno, il 1949, trovano la loro prima collocazione editoriale sia *Vom Ursprung und Ziel der Geschichte* (*Origine e senso della storia*) di Karl Jaspers sia *Meaning in History* (*Significato e fine della storia*) di Karl Löwith sia *Faith and History: A Comparison of Christian and Modern Views of History* (*Fede e storia. Studio comparato della concezione cristiana e della concezione moderna della Storia*) di Reinhold Niebuhr. Le opere, pur nelle loro differenze, testimoniano un rinnovato interesse critico verso l'interpretazione filosofica dell'origine e del senso del procedere storico in un universo – quello post-bellico – che sente la necessità di una ricostruzione etica e politica dopo le macerie del conflitto e che intende trovare in una *nuova* filosofia della storia il fondamento legittimante di un ordine internazionale pacifico.

³ K. Jaspers, *Origine e senso della storia*, cit., p. 263.

⁴ In particolar modo nel testo *Lo spirito europeo*, Jaspers riflette sull'esigenza di una conversione etico-politica dell'Europa, necessaria per ricostruire la propria identità in frantumi e per comprendere il proprio ruolo all'interno di un nuovo ordine mondiale (Cfr. K. Jaspers, *Vom europäischen Geist*, Piper, München 1946; tr. it. *Lo spirito europeo*, a cura di R. Celada Ballanti, Morcelliana, Brescia 2019).

⁵ Per un approfondimento sul tema, si veda G. Costanzo, *La nascita, inizio di tutto. Per un'etica della relazione*, Orthotes, Salerno 2018; M. Vergani, *Nascita. Una fenomenologia dell'esistenza*, Carocci, Roma 2020.

⁶ H. Arendt, K. Jaspers, *Briefswechsel, 1926-1969*, hrsg. von H. Saner e L. Köhler, Piper, München 1985; tr. it. *Carteggio 1926-1969. Filosofia e politica*, Feltrinelli, Milano 1989.

voci, che trova in *Vom Ursprung und Ziel der Geschichte* un'interessante testimonianza. All'interno di una "costellazione teoretica condivisa"⁷ la *nascita*, nelle sue diverse articolazioni biologiche, filosofiche e politiche, costituisce il *trait d'union* che permette di pensare, insieme, il concetto arendtiano di *inizio* e quello jaspersiano di *origine*. La comune esperienza del terrore nazional-socialista, che ha devastato ogni spazio di libertà privata e pubblica, ha spinto i due filosofi, pur su sentieri differenti, a voler approfondire l'evento anarchico e rivoluzionario della nascita, facendone la *conditio sine qua non* della riflessione etica e politica.

Il contributo intende approfondire proprio la "filiazione" indiretta della categoria arendtiana di *nascita* come "nuovo *inizio*" da quella che Jaspers ha definito "rivoluzione assiale", intesa come *nascita* delle civiltà universali e *inizio* della loro coscienza storica. Anche alla luce dell'attuale dibattito sull'*Axial Age* e delle più importanti riletture critiche offerte dalla sociologia comparatistica di Eisenstadt e dalla storiografia di Asman e Bellah, risulta interessante analizzare le riflessioni che il filosofo tedesco dedica alla tesi sull'assialità, all'interno di una nuova filosofia della storia che influenzerà le analisi che, qualche anno dopo, Arendt proporrà sulla natalità e sulla rivoluzione. Tale "lessico comune" permetterà di rileggere l'*origine* e il *senso* della storia delineata da Jaspers come un processo di *trasformazione*, di possibile *nascita* e *rinascita* dell'esistenza, in vista di un cammino spirituale e politico da realizzare insieme – Oriente e Occidente – verso una *nuova età assiale* dell'umanità.

1. L'assialità, l'*inizio* della storia e la *seconda* nascita della coscienza

A partire dall'esperienza del "ferreo vincolo" del terrore totalitario – che distrugge la pluralità umana "fondendola" nell'illusione violenta di un *unico* uomo – Hannah Arendt inizia a riflettere sulla questione della nascita⁸. Come scrive nelle ultime pagine di *The Origins of Totalitarianism*, anche lì dove il mondo sembra essere giunto alla sua fine

⁷ Per un approfondimento sul tema, rimando al mio saggio *Lo spettro del dominio totalitario e l'eclissi della libertà. In dialogo con K. Jaspers e H. Arendt* (Rivista Jaspersiana, XII, 2024, pp. 199-218) che inizia a tratteggiare quella che può essere definita una "riflessione a due voci" tra Arendt e Jaspers, coltivata in oltre quarant'anni di profondo rapporto personale e filosofico. Il continuo dibattito tra il Maestro e l'allieva permette, infatti, di rintracciare, pur nelle divergenze, alcuni snodi teoretici comuni, in particolar modo il concetto di "totalitarismo", con le annesse analisi sulla *massa* e sulla *tecnica*. Rispetto a tali indagini, il presente contributo intende porsi in continuità, approfondendo un'ulteriore questione di dibattito tra i due filosofi: il concetto di *nascita*.

⁸ Per un approfondimento sul tema si vedano, tra gli altri, P. Ricci Sindoni, *Hannah Arendt. Come raccontare il mondo*, Studium, Roma 1995; L. Boella, *Hannah Arendt. Agire politicamente, pensare politicamente*, Feltrinelli, Milano 2005; A. Papa, *Hannah Arendt*.

esso contiene necessariamente un nuovo inizio; questo inizio è la promessa, l'unico messaggio che la fine possa presentare. L'inizio, prima di diventare avvenimento storico, è la suprema capacità dell'uomo; politicamente si identifica con la libertà umana. «*Initium ut esset, creatus est homo*, affinché ci fosse un inizio, è stato creato l'uomo», dice Agostino. Questo inizio è garantito da ogni nuova nascita; è in verità ogni uomo.⁹

La categoria ontologica e politica dell'*initium* permette ad Arendt di pensare il "miracolo" dell'azione attraverso cui la libertà umana, interrompendo l'automatismo biologico e fratturando l'ordine esistente, dà origine al *novum*. Come scrive la filosofa in *Human Condition*: "il miracolo che preserva il mondo [...] dalla sua normale, *naturale* rovina è in definitiva il fatto della *natalità*, in cui è ontologicamente radicata la facoltà di agire. È, in altre parole, la nascita di nuovi uomini e il *nuovo inizio*, l'azione di cui essi sono capaci in virtù dell'esser nati"¹⁰. Attraverso l'azione – sempre plurale, corale, condivisa nello spazio pubblico dell'*infra* – ogni uomo si inserisce nel mondo attraverso una "*seconda nascita*"¹¹ che ha un potenziale *rivoluzionario* poiché interrompe l'esistente per *iniziare* qualcosa di radicalmente nuovo¹². "Ogni nuova nascita – scrive Arendt – è come una garanzia della salvezza del mondo, come una promessa di redenzione per chi non è più inizio"¹³. Tali riflessioni etiche e politiche

Per una filosofia della vita, Pergola, Lecce 1993; Ead., *Nati per incominciare. Vita e politica in Hannah Arendt*, Vita e Pensiero, Milano 2011.

⁹ H. Arendt, *The Origins of Totalitarianism*, Harcourt Brace Jovanovich, New York 1951; tr. it. *Le origini del totalitarismo*, Einaudi, Torino 2009, pp. 655-656. La citazione agostiniana è tratta dal *De civitate Dei*, libro 12, cap. 20. Cfr. Agostino, *La città di Dio*, a cura di Domenico Marafioti, Mondadori, Milano 2015.

¹⁰ H. Arendt, *The Human Condition*, University of Chicago Press, Chicago 1958; tr. it. *Vita activa. La condizione umana*, Giunti Editore, Firenze 2006, p. 264.

¹¹ Ivi, p. 195.

¹² Interessante in tal senso rileggere quanto Arendt pronunciò in un'intervista del 1970 in risposta ai movimenti *rivoluzionari* studenteschi e di protesta politica, tra i quali circolava il suo testo *On revolution* (1963): "l'esperienza di essere liberi coincideva, o meglio era intimamente intrecciata, [...] metaforicamente parlando, con la nascita di una nuova era. Essere liberi e dare inizio a qualcosa di nuovo erano percepiti come la stessa cosa. E ovviamente questo misterioso dono umano, la capacità di iniziare qualcosa di nuovo, ha qualcosa a che fare con il fatto che ciascuno di noi entra nel mondo come un nuovo arrivato attraverso la nascita. In altre parole, possiamo dare inizio a qualcosa perché siamo inizi e perciò iniziatori. [...]. Il significato della rivoluzione è l'attuazione di una delle più grandi potenzialità umane, l'impareggiabile esperienza di essere liberi di produrre un nuovo inizio, dalla quale viene l'orgoglio di aver aperto il mondo a un *Novus Ordo Saeculorum*" (H. Arendt, "The freedom to be free", in J. Kohn (ed. by), *Thinking without banister. Essays in Understanding (1953-1975)*, Random House, New York, 2018, p. 384). Per un approfondimento sul tema si veda anche H. Arendt, *On revolution*, The Viking Press, New York 1963; tr. it. *Sulla rivoluzione*, Einaudi, Torino 2006.

¹³ H. Arendt, *Quaderni e diari. 1950-1973*, tr. it. di C. Marazia, Neri Pozza, Vicenza 2007, p. 308.

hanno nella filosofia jaspersiana, in particolar modo nel suo concetto di *origine* e di *rivoluzione assiale*, alcune premesse problematiche che adesso si intende analizzare.

Il concetto di “età assiale” (*Achsenzeit*), inteso come *nascita* spirituale dell’“uomo, autentico”¹⁴, emerge nella riflessione di Jaspers sin dai *Grundsätze des Philosophierens* del 1942-43¹⁵, per poi trovare una trattazione più sistematica nella conferenza *Vom europäischen Geist* (1946) e in *Vom Ursprung und Ziel der Geschichte* (1949). Proprio quest’ultimo testo intende delineare una nuova filosofia della storia che, superando il modello “classico” deterministico di matrice hegel-marxiana, non rinunci al proposito di individuare un’*origine* e, di conseguenza, un *senso* della storia. Riconosciuta la necessità e il limite della conoscenza storiografica e dell’analisi empirica delle fonti, l’interpretazione jaspersiana della storia si pone kantianamente come esigenza insopprimibile della ragione. Si lega infatti ad essa:

un atteggiamento d’attesa e di ricerca della verità, di non-sapere-ancora neppure ciò che è già, ma che diverrà pienamente comprensibile soltanto dal punto di osservazione del futuro. In questo atteggiamento di fondo, persino il passato è inconcluso [...]. Ciò che fu è tuttora suscettibile di nuova interpretazione. Ciò che sembrava deciso diviene di nuovo domanda [...]. Ciò che è stato ha ancora da rivelarci che cosa è.¹⁶

Riscoprire il *desiderio* e, al contempo, *la necessità* di una nuova filosofia della storia riflette, per il filosofo tedesco, una responsabilità etica e politica verso l’umanità dopo gli atroci eventi del nazional-socialismo tedesco e del comunismo sovietico. È necessario per riscrivere una storia condivisa, *oltre* ogni particolarismo etnico, ogni nazionalismo esasperato o integralismo religioso violento, *senza* eludere le differenze culturali e sociali.

Prima di proporre la sua tesi, Jaspers afferma il suo debito teorico nei confronti di alcuni studiosi, cui riconosce il merito di aver delineato per la prima volta l’idea di “civiltà” intesa come “organismo storico unitario”: Oswald Spengler ne *Der Untergang des Abendlandes* (1918)¹⁷, Arnold Toynbee in *A study of history* (1935)¹⁸, ma soprattutto Alfred

¹⁴ K. Jaspers, *Origine e senso della storia*, cit., p. 22.

¹⁵ Cfr. Oliver Immel (Hg.), *Karl Jaspers Gesamtausgabe. Schriften zur Philosophie* (1938-1961), Schwalbe Verlag, Hannover 2024.

¹⁶ K. Jaspers, *Origine e senso della storia*, cit., p. 328.

¹⁷ O. Spengler, *Der Untergang des Abendlandes. Umriss einer Morphologie der Weltgeschichte*, C.H. Beck’sche Verlagsbuchhandlung Oskar Beck, München 1918; tr. it. *Il tramonto dell’Occidente*, Longanesi, Milano 2008.

¹⁸ A. Toynbee, *A study of history*, Oxford University Press, London 1934-1961, voll. 12; tr. it. *Le civiltà nella storia* (Compendio di D.C. Somervell), Einaudi, Torino 1950.

Weber con la sua *Kulturgeschichte als Kultursoziologie* (1935)¹⁹. A partire dalle loro riflessioni, è allora possibile cominciare a pensare quella che il filosofo tedesco definisce “rivoluzione assiale”, intesa come un punto di catastrofe nella storia universale dell’uomo. Tale mutamento paradigmatico, è rintracciabile lungo un “asse empiricamente riconoscibile” del tempo che, in senso *orizzontale* demarca una distinzione fondamentale tra la preistoria, la prima parte della storia delle civiltà millenarie e la storia mondiale moderna²⁰; ma al contempo dà origine a una *verticalità* intesa come dimensione trascendente che costituisce un nuovo *inizio* della storia inteso come *frattura* con il passato e *rielaborazione critica* di esso.

Tale asse *globale* (e non più soltanto occidentale) della storia universale, permette di rileggere il passato, il presente e il futuro in modo *alternativo*, a partire cioè dall’emergere di una “comunanza di senso”²¹, di una di “radicale metamorfosi dell’essere-umano” e “una metamorfosi totale della storia”²². Proprio quest’età di rinnovamento può essere interpretata come una *seconda* nascita, dell’uomo e della storia, poiché permette di rintracciare una “struttura comune di autocomprensione per tutti i popoli”²³. Questo “punto di svolta” o “asse della storia mondiale” (*Achse der weltgeschichte*) non può essere più indicato – come ha sostenuto la filosofia della storia di tradizione cristiana, da Agostino a Hegel²⁴ – nell’evento straordinario della *Menschwerdung*, oggetto della

¹⁹ A. Weber, *Kulturgeschichte als Kultursoziologie*, Piper, München 1935; tr. it. *Storia della cultura come sociologia della cultura*, Novecento, Palermo 1989. La prima elaborazione del concetto di “età assiale” è da ricondurre allo studioso iranologo Abraham-Hyacinth Anquetil-Dupéron (1731-1805) che osservò già nel 1771 il sincronismo tra Zoroastro, Confucio, Laotse, Buddha, i profeti di Israele e i filosofi greci; parlò di “una grande rivoluzione della specie umana”. Quanto ai “predecessori” di questo modo innovativo di interpretare la storia, Jaspers menziona Eduard Meyer (1855-1930), Ernest v. Lasaulx (1805-1861), Victor von Strauss (1809-1899) e Herman Keyeserling (1880-1946), ma soprattutto Alfred Weber (1868-1958) secondo il quale l’introduzione del cavallo come mezzo di locomozione avrebbe generato un mutamento “assiale” in tutta l’Eurasia. (Per un approfondimento sul tema si veda D. Metzler, *Kleine Schriften zur Geschichte und Religion des Arttertums und deren Nachleben*, Ugarit, Munster 2004, p. 565).

²⁰ K. Jaspers, *Origine e senso della storia*, cit., p. 40.

²¹ Ivi, p. 32.

²² Ivi, p. 181 e p. 186.

²³ Ivi, p. 19.

²⁴ Proprio Hegel nelle *Lezioni di filosofia della storia* usa l’espressione di “punto di svolta” o “cardine” costituito, nella sua visione della storia universale, dall’avvento del Cristianesimo (Cfr. G.W.F. Hegel, *Vorlesungen über die Philosophie der Geschichte*, Duncker und Humblot, Berlin 1837; tr. it. *Lezioni di filosofia della storia*, Laterza, Roma-Bari 2003). Sempre a Hegel bisogna far risalire la prima strutturata definizione di “civiltà”, intesa come spirito oggettivo di un dato popolo, che si incarna nelle sue istituzioni, nella cultura, nella lingua, nell’arte, nella filosofia, esprimendone la massima realizzazione in un dato tempo. Così scrive: “[...] la cultura è proprio l’attività di rivestire il contenuto esistente con la forma dell’universalità [...] proprio come la riflessione produce le nozioni generali

fede cristiana, ma deve poter essere valido per tutti i popoli. Esso si potrebbe allora collocare fra l'800 e il 200 a.C., in cinque differenti "luoghi di irruzione" che, in parallelo, hanno visto la *nascita* di diverse civiltà: la Grecia, la Palestina, la Persia, la Cina e l'India. Scrive Jaspers:

In questo periodo si concentrano i fatti più straordinari. In Cina vissero Confucio e Lao-tse, sorsero tutte le tendenze della filosofia cinese, meditano Mo-ti, Chuang-tse, Lieh-tsu e innumerevoli altri. In India apparvero le Upanishad, visse Buddha e, come in Cina, si esplorarono tutte le possibilità filosofiche fino allo scetticismo e al materialismo, alla sofistica e al nichilismo. In Iran Zarathustra propagò l'eccitante visione del mondo come lotta fra bene e male. In Palestina fecero la loro apparizione i profeti, da Elia a Isaia e Geremia, fino a Deutero-Isaia. La Grecia vide Omero, i filosofi Parmenide, Eraclito e Platone, i poeti tragici, Tucidide e Archimede. Tutto ciò che tali nomi implicano prese forma in quei pochi secoli quasi contemporaneamente in Cina, in India e nell'occidente, senza che alcuna di queste regioni sapesse delle altre.²⁵

La comune *nascita* delle diverse civiltà non deve essere intesa hegelianamente come una successione dialettica e gerarchica ma, come scrive Jaspers, come "un'esistenza fianco a fianco, ma senza contatto, nella stessa epoca" poiché "più strade separate all'*origine* sembrano condurre alla stessa *meta*"²⁶. Tale *origine*, simultanea e parallela, permette di pensare formazioni spirituali autonome che, pur avendo una "parentela" profonda, si articolano su sviluppi storico-culturali diversi e indipendenti, di cui è possibile delineare uno schema. Secondo Jaspers, ci sono quattro fondamentali cesure in cui "l'uomo sembra essere partito [...] da una nuova base"²⁷: il periodo preistorico; il periodo delle "antiche alte civiltà" insediate in Mesopotamia, in Egitto, nella valle dell'Indo e in Cina; l'età assiale in India, Cina e Grecia; la modernità, caratterizzata in Occidente dalla nascita della scienza e della tecnica, strettamente interconnessa con l'epoca attuale che si compie nell'odierna globalizzazione. Questi quattro *stadi* della storia mondiale, "l'età prometeica, l'epoca delle antiche alte civiltà, l'epoca

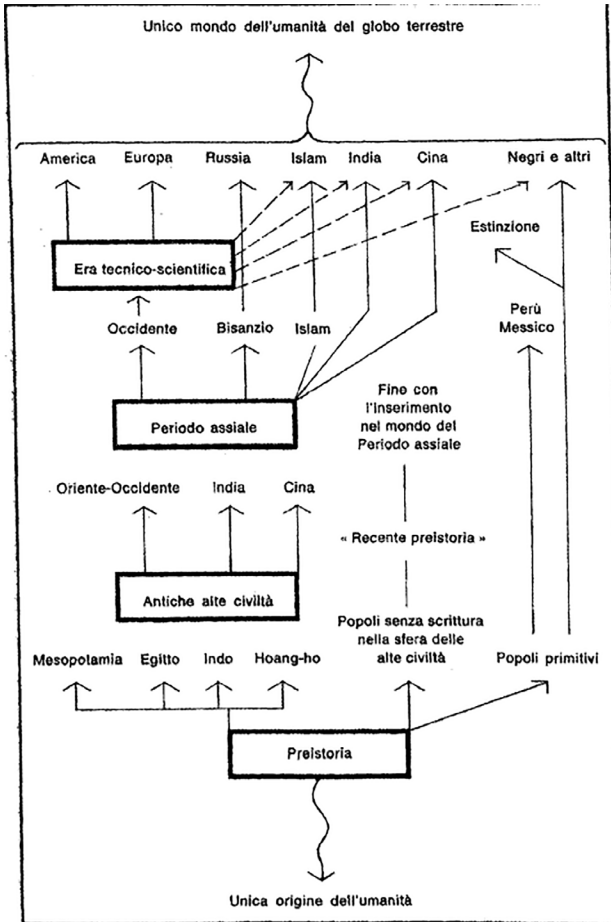
di che cosa siano genio, talento, arte, filosofia, così la cultura formale non soltanto può, anzi deve prosperare a ogni grado delle figure dello spirito e deve arrivare a una fioritura elevata. Questo grado si perfeziona fino a dare origine a uno Stato e su codesta base della civiltà esso procede alla riflessione intellettuale [...]. Nella vita dello Stato deve fare la sua apparizione anche la filosofia [...], in questo modo la filosofia, la quale è coscienza di questa forma, pensiero del pensiero, trova già preparato nella cultura generale il materiale adatto al suo edificio" (Ivi, p. 61).

²⁵ K. Jaspers, *Origine e senso della storia*, cit., p. 20.

²⁶ Ivi, p. 30, cors. mio.

²⁷ Ivi, p. 46.

della fondazione spirituale del nostro essere-umano così com'è oggi, e l'era della tecnica²⁸ sono eterogenei e possono essere ben visualizzati a partire dallo schema offerto in *Vom Ursprung und Ziel der Geschichte*²⁹.



È interessante notare, come scrive Jaspers, che ciò che definiamo “storia” ha avuto *inizio* circa cinquemila anni fa. Se la preistoria, in quanto “fiume di continui mutamenti”, è più simile al corso *naturale e biologico* degli eventi, in modo diverso, “esiste storia soltanto dove esiste anche una *conoscenza* della storia. [...]. La storia è quella frazione del passa-

²⁸ *Ibidem.*

²⁹ Cfr. *ivi*, p. 48.

to che in un dato momento è chiaramente *visibile* all'uomo"³⁰. Storia, dunque, è propriamente *coscienza di essere storico*, appropriazione di contenuti di natura *spirituale e tecnica*, nettamente differenti dal *naturale* procedere biologico, dagli impulsi vitali e dagli istinti elementari che si sviluppano nel periodo preistorico. Quest'ultimo costituisce la condizione fondamentale di possibilità dell'avviarsi della *svolta* storica, segnata dall'invenzione della scrittura, dal sorgere del linguaggio. È in questo momento che l'esistenza dell'uomo, da mero processo biologico, ha *iniziato* ad essere *storia*: "attraverso la storia l'uomo è diventato l'essere che si sforza di elevarsi al di sopra di se stesso", ha preso coscienza di sé e della propria *transitorietà*³¹.

Quel che noi chiamiamo storia, e che nel vecchio senso è ora alla fine, è stato un momento intermedio di cinquemila anni fra il popolamento del globo, svoltosi in centinaia di migliaia d'anni di preistoria, e l'odierno inizio dell'autentica storia universale. Prima della storia, tutto ciò che avvenne nell'isolamento dei gruppi umani ignari della loro interdipendenza fu un continuare a vivere interamente ripetitivo, ancora strettamente legato alle vicende della natura. Poi, la nostra breve storia è stata come un incontrarsi, un radunarsi di uomini per l'azione della storia mondiale, è stata l'acquisizione spirituale e tecnica dell'equipaggiamento necessario per affrontare il viaggio. Cominciamo giusto adesso.³²

Con un'immagine significativa, Jaspers descrive il lungo percorso della storia umana come "due lunghi respiri", di cui il primo sarebbe durato "dall'età prometeica attraverso le antiche alte civiltà fino al periodo assiale e alle sue conseguenze", mentre il secondo sarebbe iniziato con la "nuova era prometeica", ossia l'era tecnico-scientifica moderna, e potrebbe condurre "ad un nuovo periodo assiale di umanizzazione autentica, ancora remoto e a noi invisibile"³³. Mentre il *primo* respiro è "frazionato in tanti respiri paralleli, il secondo è tratto dall'intera umanità". Durante il primo respiro, ogni avvenimento è stato *locale*, una varietà di eventi tra loro isolati. In modo differente, il "balzo in avanti" costituito dall'età assiale, che dà avvio al *secondo* respiro, permette di leggere il passato a partire da un punto di vista non più soltanto particolare, ma *universale*³⁴.

Per questo la svolta assiale può esser interpretata come un *risveglio* spirituale, in quanto costituisce, secondo Jaspers, la *nascita* spirituale della coscienza, una "seconda" nascita, dopo l'inizio storico delle antiche

³⁰ Ivi, p. 49, cors. mio.

³¹ Ivi, p. 73.

³² Ivi, p. 45.

³³ Ivi, p. 46.

³⁴ Cfr. ivi, p. 102.

alte civiltà, che le permette finalmente di comprendersi all'interno di una storia *mondiale e cosmopolita*. Elemento comune di questo periodo di *rivoluzione* assiale è che “i problemi capitali si presentano nelle situazioni-limite umane – che l'uomo si conosce in tutta la sua fragilità”³⁵. Come argomenta il filosofo: “l'uomo prende coscienza dell'essere nella sua interezza, di se stesso e dei suoi limiti [...]. Pone domande radicali [...]. Incontra l'assolutezza nella profondità di essere se stesso e nella chiarezza della trascendenza”³⁶.

In questo periodo vi fu una linea di demarcazione radicale, “vennero formulate le categorie fondamentali, secondo cui pensiamo ancor oggi, e poste le basi delle religioni universali di cui vivono gli uomini”³⁷, si iniziò il processo di distanziamento dal mito e di razionalizzazione del mondo. La simbolizzazione assiale coincise dunque con un processo di distacco critico dall'immediatezza, di “*differenziazione*”, dunque di “*spiritualizzazione*” (*Vergeistigung*)³⁸. Ha marcato il passaggio dal politeismo al monoteismo, l'apertura alla Trascendenza, la scoperta di una tensione verticale che struttura il modo *umano* di abitare il mondo.

Ha costituito quella che Eisenstadt definisce una “*rivoluzione epistemologica*” in cui non più il *mythos* ma il *logos* diventa la base filosofica e teologica di elaborazione di *nuove* “immagini del mondo”, di nuove “mappe cognitive” con cui interpretare l'essere umano e l'universo in cui si muove³⁹. Questa frattura storico-spirituale provocò la *nascita* di una coscienza pienamente teoretica e per la prima volta aperta alla Trascendenza: “in ogni senso fu compiuto il passo nell'universale”⁴⁰.

Jaspers raffigura metaforicamente l'età assiale come un “*respiro*”⁴¹, un “*fermento*”⁴², una “*seconda nascita*”, un “*risveglio spirituale*”, una “*iniziazione dell'essere-umano*”⁴³, caratterizzata da un iniziale fuga dal mondo che fu fondamentale all'uomo per assumerne una distanza critica, esaltare la ragione e far emergere la coscienza dell'unicità dell'esistenza sulla terra. La rivoluzione assiale è caratterizzata, dunque, dalla consapevolezza di poter fuoriuscire dalla ciclicità naturale per trasformare l'esistenza collettiva a partire dall'azione umana.

³⁵ K. Jaspers, *Lo spirito europeo*, cit., p. 27.

³⁶ K. Jaspers, *Origine e senso della storia*, cit., p. 20.

³⁷ Ivi, pp. 19-21.

³⁸ Per un approfondimento sul tema si veda anche E. Voegelin, *Order and History*, vol. IV, *The Ecumenic Age*, University of Missouri, Columbia 2000.

³⁹ Cfr sul tema S.N. Eisenstadt, *The Origins and Diversity of Axial Age Civilizations*, New York University Press, New York 1986.

⁴⁰ K. Jaspers, *Origine e senso della storia*, cit., p. 21.

⁴¹ Ivi, p. 46.

⁴² Ivi, p. 77.

⁴³ Ivi, pp. 78-82.

Finalmente consapevole di poter mutare il proprio destino, l'uomo abita in modo nuovo il presente, in cui può "incominciare qualcosa di straordinario"⁴⁴.

In questi passaggi della riflessione jaspersiana non si può non percepire l'eco di quanto verrà poi teorizzato da Arendt a partire da *The Origins of Totalitarianism* (1951) e in particolar modo in *Human Condition* (1958) sulla nascita, sull'azione, sulla capacità dell'uomo di essere nuovo inizio. Proprio il concetto di un'Origine intesa come *inizio, senso e meta* dell'agire, mai del tutto data, ma sempre da realizzarsi attraverso un percorso di *chiarificazione* esistenziale e di *risveglio* spirituale istituisce il "balzo in avanti" dell'età assiale, in cui per la prima volta l'uomo – potremmo dire, insieme ad Arendt – si comprende come "nuovo venuto e iniziatore grazie alla nascita"⁴⁵.

La descrizione della storia come *inizio* della coscienza storica e, in modo più specifico, della rivoluzione assiale come *seconda* nascita della coscienza permettono a Jaspers di rinnovare la *fede filosofica* in un'unica *origine* dell'umanità⁴⁶. Ma "l'unità [della storia] non è un fatto bensì un

⁴⁴ Ivi, p. 23.

⁴⁵ H. Arendt, *Vita activa*, cit., p. 195.

⁴⁶ Il concetto jaspersiano di "fede filosofica" si riferisce a quell'apertura strutturale dell'esistenza alla Trascendenza che, senza presumere di auto-costituirsi come sapere di una verità ultima, si mantiene costantemente in bilico di fronte al rischio, alla scommessa, alla problematicità. Pur condividendo con la fede *religiosa* la convinzione che la verità totale non si realizzi nel tempo, essa se ne distanzia poiché non pretende di poter acquisire una verità universalmente valida: "la fede filosofica non scomunica, non dichiara eretici i dissenzienti, non accende roghi, non dispone di libri sacri privilegiati rispetto ad altri, perché contenenti la verità assoluta; la fede filosofica sa di essere per via (*auf dem Wege der Wahrheit*) e non scambia la via incerta con la meta definitiva" (U. Galimberti, *Introduzione a K. Jaspers, Der philosophische Glaube*, Piper Verlag, München 1948; tr. it. *La fede filosofica*, Raffaello Cortina, Milano 2005, p. 41). Oltre l'alternativa tra fede rivelata, scienza totale e nichilismo, la fede filosofica jaspersiana preferisce rimanere *sempre* aperta all'*Umgreifende*, il "Tutto-abbracciante" (Cfr. K. Jaspers, *La fede filosofica*, cit., in particolar modo pp. 70-71). Riprendendo queste tematiche, Jaspers scrive in *Origine e senso della storia*: "La fede è il tutto-abbracciante che è al timone anche quando l'intelletto sembra fare per conto proprio. Fede non significa un determinato contenuto, o un dogma; il dogma può essere espressione di una forma storica assunta dalla fede, ma può anche ingannare. La fede è l'elemento integratore e motore nel fondo dell'uomo, in cui l'uomo è legato, al di sopra e al di là di se stesso, all'origine dell'essere. L'evidenza della fede si ottiene soltanto in forme storiche; nessuna verità può considerarsi l'unica ed esclusiva per tutti gli uomini senza diventare intollerante e allo stesso tempo non vera; ma fra tutti i credenti esiste un elemento comune nascosto. [...] L'uomo non è semplicemente una creatura istintiva o un punto d'intelligenza, bensì una creatura che è per così dire al di sopra e al di là di se stessa. Non è mai esaurito da ciò che diviene come oggetto della fisiologia, della psicologia e della sociologia. Può prender parte a un tutto-abbracciante, attraverso il quale soltanto diviene autenticamente se stesso" (Id., *Origine e senso della storia*, cit., pp. 272-273).

fine⁴⁷, non può mai esser completata – “è un tutto aperto” alla libertà dell'uomo⁴⁸. L'unità può essere solo simbolizzata e perseguita attivamente come “compito infinito”⁴⁹ come unità di *senso* da costruire insieme, oltre le barriere storico-culturali tradizionali.

La situazione oggi esige che ritorniamo a una più profonda origine, alla fonte da cui ogni fede un tempo è scaturita nelle sue particolari forme storiche, a quella fonte che può sgorgare ogni volta che l'uomo è pronto per essa. Quando la fiducia in ciò che appare ed è dato nel mondo non sostiene più la vita, è la fiducia nell'origine di tutto che deve porre le fondamenta.⁵⁰

In questa *fede* nella libertà, in questa “fiducia dell'origine”, chiamata a realizzarsi nel futuro è possibile intravedere un'*idea* regolativa della ragione o, con le parole di Jaspers, “un'unità spirituale indefinitamente mobile e aperta, in nome di quell'Origine che sempre ci accompagna”⁵¹. Solo a partire dalla rinnovata situazione *spirituale* del nostro tempo, nell'attuale angoscia per il futuro dell'umanità di fronte alla nuova “situazione-limite”⁵² della bomba atomica, può nascere un vero interesse per il periodo assiale. È infatti solo *ora*, che a seguito della compiuta unificazione del globo terrestre, si avverte il bisogno di indagare le grandi tradizioni culturali da cui l'uomo proviene. Solo *adesso* si pone il problema non delle *cause* ma del *senso* dell'età assiale, intesa come esigenza di “acquisire qualcosa di comune a tutta l'umanità al di sopra di ogni differenza di credo”, dunque di “invito alla comunicazione *illimitata*”⁵³, come richiamo al una conversione etico-politica, necessaria per far fronte alle sfide del presente. “Ecco, in questa prospettiva, il problema di ognuno: dove prender posizione, per che cosa operare”⁵⁴.

C'è da chiedersi: come si preserverà, nelle condizioni dell'era della tecnica e del riordinamento dell'intera comunità umana, il contenuto del patrimonio tramandato: l'infinito valore dell'individuo, la dignità e i diritti umani, la libertà dello spirito, le esperienze metafisiche di millenni? Ma l'autentico interrogativo del futuro, che tutto condiziona e include, concerne come e in che cosa l'uomo crederà.⁵⁵

⁴⁷ Ivi, p. 289.

⁴⁸ Ivi, pp. 339-340.

⁴⁹ Cfr. ivi, pp. 297-298.

⁵⁰ Ivi, p. 271.

⁵¹ *Ibidem*.

⁵² Cfr. sul tema K. Jaspers, *Die Atombombe und die Zukunft des Menschen*, München, Piper, 1958; tr. it. *La bomba atomica e il destino dell'uomo*, PGRECO, Milano 2013.

⁵³ K. Jaspers, *Origine e senso della storia*, cit., p. 40.

⁵⁴ Ivi, p. 350.

⁵⁵ Ivi, p. 243.

Erede della lezione kantiana, Jaspers rinnova l'antico interrogativo sull'uomo: che cosa possiamo conoscere? Come dobbiamo agire? In cosa possiamo sperare?⁵⁶ La memoria dell'età assiale può e deve costituire uno stimolo al presente per superare pretese di assolutezza, fondamentalismo ed esclusività dei punti di vista politici, culturali e religiosi in nome della consapevolezza di esistenza di plurali "sviluppi originari dell'umanità, dotati di creazioni spirituali uniche nel loro genere"⁵⁷.

Collegata dunque all'avvenire dell'umanità, l'ipotesi jaspersiana dell'*Achszeit* risponde all'esigenza storico-filosofica della ricerca di un'*unità* nella *pluralità*, traducendosi nel tentativo di gettare un ponte tra Oriente e Occidente, nell'impegno etico-politico che vuole favorire l'*apertura* tra le culture e le religioni mondiali per rinvenire, nel libero confronto, una *comune* – seppur plurale e diversificata – Origine *condivisa*⁵⁸.

2. *The Axial Age Civilizations: un conflitto di interpretazioni*

Il percorso di autocomprensione di *tutti* i popoli, secondo binari *diversi* e, al contempo *paralleli*, cui l'età assiale permette di pensare, costituisce un'intrigante sfida teoretica e politica per il nostro presente e, indubbiamente, per il nostro futuro. Come è emerso nel famoso convegno di Erfurt nel 2008, dedicato alla memoria di Karl Jaspers, le cui riflessioni critiche hanno dato vita al volume curato da Robert Bellah e Hans Joas *The Axial Age and Its Consequences* (2012)⁵⁹, sia che si intenda l'età assiale come acquisizione storicamente verificata, sia che la si intenda come progetto ancora *in fieri*, essa continua a costituire un concetto fecondo nel dibattito contemporaneo. Il fascino esercitato dalla tesi dell'assialità su molti studiosi deriva, infatti, principalmente dalla sua capacità performativa di riconfigurare narrativamente la storia della coscienza umana – la *nascita*

⁵⁶ Cfr. I. Kant, *Kritik der reinen Vernunft*, Riga, Joh. Fr. Hartknoch, 1781; tr. it. *Critica della ragion pura*, Laterza, Roma-Bari, 1991, in particolare pp. 495 e sgg.

⁵⁷ K. Jaspers, *Lo spirito europeo*, cit., p. 27.

⁵⁸ Il significato del fine della storia, come del suo senso, presuppone, nel testo jaspersiano, l'evidente richiamo al tema dell'origine che aveva segnato già alla fine degli anni '20 il *Dramma barocco tedesco* di Benjamin, così come le riflessioni di Heidegger: "L'origine, benché sia una categoria pienamente storica, non ha nulla in comune con la genesi. Con origine non s'intende un divenire del già nato, bensì un divenire e un trapassare di ciò che nasce. L'origine sta nel fiume del divenire come un vortice e trascina dentro la propria ritmica il materiale della nascita" (W. Benjamin, *Ursprung des deutschen Trauerspiels*, Verlag, Berlin 1928; tr. it. *Il dramma barocco tedesco*, Einaudi, Torino 1971, p. 28). Interessante evidenziare già in questo scritto di Benjamin quello stretto legame tra *origine* e *nascita* che costituirà poi l'oggetto di discussione di Arendt e Jaspers.

⁵⁹ R. Bellah, H. Joas (ed. by), *The Axial Age and Its Consequences*, Harvard University Press, Cambridge 2012.

dell'esistenza “così come la conosciamo”, con le parole di Jaspers – all'interno di una prospettiva *globale* che, soprattutto oggi, è fondamentale per affrontare le nuove sfide etico-politiche di un tempo di incertezza.

Le riletture, anche critico-decostruttive, che ne sono state offerte dal dibattito filosofico, storico e sociologico attuale intendono approfondire le potenzialità e i rischi del concetto che – come ricorda Robert Bellah – rappresenta un'eredità culturale dalle potenzialità esplosive, sia nel bene che nel male⁶⁰. Esso, infatti, non è soltanto una “risorsa” per affrontare l'attuale crisi dell'uomo ma anche un “peso”. La svolta assiale è radicata in una profonda crisi di legittimazione degli ordini socio-politici pre-esistenti, rispetto ai quali si sviluppa un atteggiamento di aperto rifiuto e negazione. Proprio tale *distacco* dalla realtà esistente, secondo Bellah, è ciò che permette quel processo critico di simbolizzazione che sfocia nella *theoria*, quella che in altra sede Arendt avrebbe definito “vita contemplativa”, in contrasto con la “vita *activa*”⁶¹. Il rifiuto di un agire *immediato* nel mondo, l'esigenza di volerlo *spiegare* in modo ultimo e definitivo, sussumendo il particolare all'interno dell'universale si è tradotto in molti casi in quelli che Sloterdijk ha definito “un mondo di separatisti”⁶², in un atteggiamento rinunciatario nei confronti del mondo circostante (e rivolto alla ricerca della salvezza, individuale e collettiva, in un mondo oltre-mondano) o, in modo polarmente contrapposto, in una violenta volontà di riforma che, per imporre un *modello* alla realtà finisce per produrre una “conoscenza distaccata” che non critica, ma legittima l'ingiustizia e la repressione violenta di tutto ciò che è *diverso* rispetto alla proiezione ideale⁶³. La capacità di *distanziarsi* dalla realtà, per contestarne i limiti e i difetti, permette di formulare una critica corrosiva nei confronti dell'ordine esistente e può trasformarsi, in una “proiezione utopica di una *buona* società”, adesso inesistente ma da realizzare *con ogni mezzo*, la quale, può avere però effetti nocivi e violenti di fronte ai quali è sempre necessario rimanere vigili.

Anche Jan Assmann mette in luce alcuni rischi impliciti della teoria dell'assialità⁶⁴. Il concetto di età assiale, infatti, è il frutto di una ricerca

⁶⁰ R. Bellah, *L'eredità dell'età assiale. Una risorsa o un peso?*, tr. it. di Raffaella Cantillo, in M. Deodati, F. Miano, S. Wagner (a cura di), *Storia e età assiale. Studi Jaspersiani III*, Orthotes, Salerno 2015, pp. 73-94, in part. p. 94. Per un approfondimento sul tema, si veda anche R. Bellah, *Religion in Human Evolution: From the Paleolithic to the Axial Age*, Harvard University Press, Cambridge 2011.

⁶¹ Cfr. H. Arendt, *Vita activa*, cit.

⁶² P. Sloterdijk, *Du mußt dein Leben ändern. Über Anthropotechnik*, Suhrkamp, Frankfurt am Main 2009; tr. it. *Devi cambiare la tua vita. Sull'antropotecnica*, Raffaello Cortina, Milano 2010.

⁶³ Cfr. R. Bellah, *L'eredità dell'età assiale. Una risorsa o un peso?*, cit., pp. 94 e sgg.

⁶⁴ J. Assmann, *La memoria culturale e il mito dell'età assiale*, tr. it di Raffaella Cantillo, in M. Deodati, F. Miano, S. Wagner (a cura di), *Storia e età assiale. Studi Jaspersiani III*, cit., pp. 29-71.

volta a scandagliare le radici della modernità: nasce dunque da un'*intenzionalità* filosofica e sociologica molto forte e profondamente orientata a ritrovare nel passato alcune delle condizioni di possibilità di sviluppo delle società attuali e, soprattutto, le *cause* di alcune *degenerazioni* della modernità. Proprio tale quadro epistemologico – che è quello di una *filosofia* della storia, più che di una *storia* della filosofia – conduce, secondo Assmann, molti studiosi a sottovalutare l'effettività delle fonti storiche e filologiche, costruendo il *mito* di un'unica origine dell'uomo, dell'*unitario* punto di svolta nell'evoluzione culturale dell'umanità, necessario per legittimare il successivo *mito* della modernità come progetto ancora incompiuto. In modo differente, se l'età assiale è intesa come una "rivoluzione" che ristrutturata l'orientamento nel tempo, essa deve trovare il proprio *humus* – secondo Assmann – nella costruzione, lenta e progressiva per ogni popolo, di una "memoria culturale" condivisa e di una "riflessività discorsiva"⁶⁵. La questione decisiva, infatti, non è tanto *ciò che accadde* durante questo periodo di svolta, quanto *come* quegli eventi siano stati ricordati, rappresentati, simbolizzati e ricostruiti come patrimonio identitario nelle varie tradizioni culturali. Eventi che non solo non sono stati sincronici ma, soprattutto, hanno avuto un importante preludio nell'impero egizio – non considerato da nessun teorico dell'assialità – grazie alla diffusione della scrittura, all'alfabetizzazione e, in particolar modo della "canonizzazione", intesa come processo sociale e collettivo di costruzione di una "continuità *testuale*" che costituisce fondamento a partire dal quale si strutturano le identità sociali e culturali dei diversi popoli. Proprio tali elementi – alfabetizzazione e canonizzazione – possono diventare strumenti empirici validi per uno studio comparato delle culture che, però, *storicamente* deve rinunciare alla pretesa *filosofica* di trovarne un'origine comune e condivisa.

A partire da tali premesse critiche e dalla costante vigilanza contro il pericolo di un nuovo – e violento – *mito fondativo*, secondo Assmann, il concetto di età assiale può avere anche una peculiare e positiva funzione *euristica* nello studio della storia dei differenti popoli: è infatti possibile ritrovare nelle trame dei loro racconti – dunque nella loro memoria culturale condivisa – alcune linee di continuità. In particolar modo, le tradizioni sociali e politiche – pur nella loro incredibile diversità ed eterogeneità – narrano sempre di un momento peculiare della storia identitaria del popolo, un momento di *svolta*, caratterizzato da un duplice processo di "distanziamento" dall'adesione *immediata* alla tradizione culturale precedente – dunque un percorso di *differenziazione*, negazione e opposizione rispetto all'ordine mondano esistente – e di contemporaneo emergere di

⁶⁵ Cfr. *ivi*, pp. 50 e sgg.

una spiccata “riflessività critica”. Tale percorso critico di *spiritualizzazione* culminò nella costruzione di un ordine *ideale* (ultra-mondano) inteso come fonte legittimante dell’azione religiosa, morale e politica. Ma questo doppio percorso di *distanziamento* e *universalizzazione* non fu un orientamento generale: solo i “grandi individui” – i saggi, gli acculturati, i profeti – solo una ristrettissima *élite* fu in grado di *fare un passo indietro* rispetto alla realtà per guardare *oltre* essa. Tale “conquista” del *distanziamento* sociale e ideologico, con la conseguente capacità di messa in questione radicale e negazione dell’esistente fu peculiarità solo di pochi e, differentemente da quanto teorizzato dagli studiosi dell’assialità, non fu in grado di generare un cambiamento *permanente* negli assetti sociali, legato invece alla concomitanza storico-politica di altri – e ben più importanti – fattori di evoluzione e di sviluppo. Nonostante queste profonde critiche ad alcuni dei pilastri epistemologici della teoria assiale, Assmann riconosce però, che essa non riguarda tanto l’uomo “come lo conosciamo”, e la sua prima apparizione nel tempo, ma “l’uomo come lo vogliamo”, l’obiettivo utopico di una comunità universale civilizzata”⁶⁶. Intravedendo, dunque, nell’assialità non il valore *descrittivo* di una realtà *storicamente* data, quanto quello che Kant avrebbe definito un *ideale regolativo* della ragione, è possibile evitarne una mitizzazione e, allo stesso tempo, valorizzarne gli elementi più creativi e costruttivi, di *critica* dell’esistente e di costante *desiderio* di fioritura di un’esistenza pienamente realizzata.

Particolarmente interessante è adesso riflettere sull’originale rielaborazione della teoria dell’assialità che Samuel Eisenstadt sviluppa a partire dal 1983, in un convegno sul problema dell’origine e della diversità delle civiltà assiali organizzato a Bad Homburg, in Germania, e poi confluito in *The Origins and Diversity of Axial Age Civilizations*⁶⁷ e in *Multiple Modernities*⁶⁸.

Riprendendo l’idea jaspersiana di un asse *universale* della storia lungo il quale si realizza – pur a velocità e ritmi diversi – lo sviluppo delle civiltà, Eisenstadt individua un momento di *svolta* fondamentale, quella che weberianamente si può definire una fase di *disincantamento del mondo*, una vera e propria “rivoluzione” nell’ambito delle idee e delle strutture istituzionali delle antiche società, destinata ad avere effetti irreversibili nella storia dell’umanità. Essa chiude definitivamente con le concezioni mitiche fondate su una visione etnica e particolaristica dell’umanità, per costruire un’immagine *universale* e *razionale* dell’uomo e della coscienza.

⁶⁶ Ivi, p. 71.

⁶⁷ Cfr. S.N. Eisenstadt (ed. by), *The Origins and Diversity of Axial Age Civilizations*, cit.; J.P. Arnason, S.N. Eisenstadt, B. Wittrock, *Axial Civilizations and World History*, Brill, Boston 2005.

⁶⁸ S.N. Eisenstadt, *Comparative Civilizations and Multiple modernities*, Brill, Boston 2003.

za umana. Tali profondi cambiamenti teorici, culturali, politici e sociali sono legati all'emergere di una *distinzione* – sino ad allora pressoché assente, o ritenuta poco significativa – tra il mondo *umano* e il mondo *oltre-umano*, la nascita di una “*tensione* essenziale” tra due diversi ordini di senso dell'esistenza, quelli lo studioso definisce l'ordine “mondano” e l'ordine “trascendentale”⁶⁹. Come abitare tale tensione, come colmare questo divario? Attraverso un lungo processo di rivoluzione *culturale, sociale e politica*, orientata a trasformare l'ordine “mondano” sulla base di quello “oltre-mondano”: è proprio il processo di istituzionalizzazione – mai definitiva – di questa tensione a fornire le possibilità di una profonda critica dell'esistente. Essa avviene, in modi differenti, nell'antico Israele, nell'antica Grecia, nella prima Cristianità, nello zoroastrismo, nella Cina imperiale, nelle civiltà Hindu e buddhiste e, con modalità peculiari, anche con la nascita dell'Islam. Si caratterizza per l'emergere di nuove *élites* intellettuali, dotate di autonomia culturale e politica che, in contrasto con le gerarchie di poteri precedenti, si fanno portatrici di un'esigenza di rinnovamento, della consapevolezza di poter *cambiare* il mondo umano, poterlo ri-costruire *attivamente* a partire da un ordine “trascendentale” che lo precede e lo legittima. Tale rivoluzione dell'ordine *ideale* diventa duratura poiché si istituzionalizza nell'ordine *reale*, attraverso la costruzione di nuovi centri di potere politico e nuovi strumenti di legittimazione che permettono quella svolta sociale, politica e culturale che caratterizza le civiltà assiali. È allora possibile studiare le *molteplici* “storie del mondo” attraverso un'analisi comparativa che metta in luce le condizioni di emergenza e di istituzionalizzazione di questo periodo di rivoluzione, le modalità con cui le *antiche* organizzazioni tribali si trasformano nelle *nuove* società, le caratteristiche delle *élites* promotrici di questo cambiamento, le *visioni trascendentali* che legittimano simbolicamente questa rivoluzione politica, la *nuova* “riflessività sociale” che, anche nel conflitto violento, permette di costruire identità sociali più strutturate e complesse. A partire da questo studio comparato, sarà possibile distinguere alcune civiltà – come quella cinese antica, fondata sul confucianesimo, o quella greca – che “risolvono” la *tensione* assiale dando priorità all'ordine mondano, e altre civiltà – come l'antico Israele, la cristianità, l'Islam e, in modo differente, l'induismo e il buddhismo – che concettualizzeranno tale tensione come *iato* religioso, privilegiando infine l'ordine oltre-mondano.

Analizzare il rapporto tra le *condizioni* – sociali, culturali e politiche – di emergenza dei processi di trasformazione assiale e l'ordine *istituzionale* e *simbolico* in cui essi si realizzano permette non solo di tratteggiare la nascita delle *diverse* civiltà assiali ma anche il loro *diverso* sviluppo,

⁶⁹ Cfr. S.N. Eisenstadt, *The Axial Age Breakthroughs. Their Characteristics and Origins*, in Id., *The Origins and Diversity of Axial Age Civilizations*, cit., pp. 1-25.

la conflittualità, politica e sociale, con cui esse concretamente costruiscono la propria identità, il *diverso* processo di modernizzazione che ne ha configurato le caratteristiche attuali. Oltre l'omogeneità e l'egemonia strutturale del "modello" occidentale di modernità, è così possibile pensare "modernità multiple", molteplici differenti civiltà, divergenti ma in continua interferenza reciproca⁷⁰. Modernizzazione, infatti, non significa necessariamente "occidentalizzazione", ma è possibile pensare – su suggerimento di Eisenstadt – diversi sistemi moderni, non diretti da un centro, ma pluralistici e pluricentrici, guidato da dinamiche autonome. Esistono dunque diversi modi di rispondere alla "sfida" occidentale della modernizzazione, "diversi modi di essere moderni", tutti radicati nella *nascita* della consapevolezza di poter *iniziare* qualcosa di nuovo, in modo indipendente rispetto al passato, dalla certezza che la società può essere *rinnovata* dall'attività dell'uomo, pur all'interno di un processo conflittuale di trasformazione e sviluppo⁷¹.

Jaspers si pose esplicitamente la domanda se l'età moderna e contemporanea potessero rappresentare una nuova civiltà assiale, ma la questione rimase allora problematica: secondo il filosofo tedesco l'attuale epoca della globalizzazione avrebbe potuto dar vita a una seconda *rivoluzione* assiale solo dopo una profonda *rivoluzione spirituale*, del singolo e della collettività. La modernità infatti, a differenza dell'età assiale, non è un momento di rinascita *spirituale*, piuttosto uno sviluppo *materiale*, un nuovo "prometeismo" della tecnica che si realizza nell'invenzione di nuovi utensili e nuove armi, "un'epoca di reale rimodellamento tecnico e politico, non un'epoca di creazioni spirituali eterne"⁷².

In modo differente, Eisenstadt sostiene che la radicale trasformazione rappresentata dalla modernità, la sua apertura alla possibile unificazione globale, costituisce a tutti gli effetti un *nuovo* periodo assiale, chiamato a fronteggiare nuove antinomie e paradossi⁷³. Gli attuali processi di globalizzazione non comportano – secondo lo studioso – né la "fine della storia", nel senso di una fine degli scontri ideologici tra i diversi programmi culturali della modernità, né uno "scontro di civiltà", che impegna un Occidente secolare nel confronto con società che non intendono aderire al programma della modernità. Piuttosto esse

⁷⁰ Cfr. S.N. Eisenstadt, *Comparative Civilizations and Multiple modernities*, cit.

⁷¹ Cfr. *ivi*, pp.5 e sgg.

⁷² K. Jaspers, *Origine e senso della storia*, cit., p. 182.

⁷³ Per un approfondimento critico si legga R. Regni, *Introduzione* a K. Jaspers, *Origine e senso della storia*, cit., pp. XII-XV, in part. p. XIV in cui scrive: "Se questa [modernità] vuol dire uso della tecnologia, una economia di mercato, una certa miscela di uguaglianza e partecipazione, di libertà e autorità, una forma di stato-nazione o stato territoriale, allora Europa, Stati Uniti, Cina India, Giappone, paesi islamici e ed africani mostrano modi diversi di essere moderni".

propongono una “continua reinterpretazione del programma culturale della modernità”, tentando di ridefinirne il discorso in *nuovi* termini, attraverso *nuove* interrogazioni e in *nuove* arene di dibattito, oltre il primato del tradizionale Stato-nazione⁷⁴.

Le tesi sull'assialità costituiscono dunque una “sfida teorica”⁷⁵, soprattutto per quell'Occidente *prigioniero* di se stesso⁷⁶, che è costretto oggi a metter in discussione il proprio primato culturale, politico ed economico, la violenza con cui ha per secoli oppresso e colonizzato il mondo. L'idea di età assiale – come suggerito da Assmann – rivela non solo un aspetto *empirico-descrittivo* – che, come si è sinora visto, è stato profondamente criticato e decostruito – ma soprattutto una *normatività* etica e politica ancora significativa, poiché delinea un *ideale* regolativo che può guidare la ragione in un continuo oltrepassamento di sé. Tale *ideale* può concretamente realizzarsi all'interno di un nuovo ordine internazionale pacifico. La filosofia della storia proposta da Karl Jaspers, gravitante attorno all'idea di *Axial Age*, è infatti orientata a costituire il fondamento filosofico legittimante di un *nuovo* ordine mondiale che, superando l'equilibrio del terrore tra i due blocchi della guerra fredda, sia in grado di garantire una *nuova* cittadinanza cosmopolitica. Memore della lezione kantiana, secondo Jaspers è infatti possibile pensare il “cittadino del mondo”, colui che, oltre ogni appartenenza a uno Stato-nazione, gode del diritto di cittadinanza. È proprio questa ricaduta giuridico-politica a costituire l'autentica forza della riflessione jaspersiana: oltre l'illusione di potere *fondare* filosoficamente la politica, la riflessione dello studioso tedesco intende offrire l'intelaiatura filosofica per un rinnovato sistema dei diritti, in vista di un *nuovo* cosmopolitismo. Proprio quest'ultimo trova il proprio radicamento in un nuovo “*ethos* di vita comunitaria”, che si realizza in una democrazia deliberativa, consapevole delle proprie fragilità e aperta nonostante tutto al compito della “comunicazione sconfinata”⁷⁷.

La “svolta assiale” dunque, secondo Jaspers, pur avendo costituito un evento culturale “concluso” nel tempo e nello spazio, può dar vita a un processo di costante *rinascita* spirituale dell'uomo e dei differenti popoli, a partire dal quale è possibile delineare il progetto di una storia della filosofia mondiale (*Weltphilosophie*), che si intende adesso approfondire.

⁷⁴ Cfr. S.N. Eisenstadt, *Comparative Civilizations and Multiple modernities*, cit., pp. 23-24.

⁷⁵ Cfr. P. Costa, *La sfida teorica dell'assialità*, “Politiche&Società”, 2, 2015, pp. 169-190. Per un approfondimento sul tema si veda inoltre V. Cotesta, *La terza età assiale. Alcune considerazioni sulla nuova forma del mondo*, “SocietàMutamentoPolitica”, 12, 23, 2021, pp. 185-198.

⁷⁶ Cfr. M. Kundera, *Un Occidente prigioniero*, Adelphi, Milano 2022.

⁷⁷ Cfr. K. Jaspers, *Origine e senso della storia*, cit., p. 211.

3. Per una storia *mondiale* della filosofia

Grazie agli studi sulle culture indiane e cinesi consigliate dall'amico Zimmer, in Jaspers, in esilio a Ginevra, sin dal 1937 nasce il proposito di scrivere una storia universale della filosofia. Se infatti il totalitarismo era stato il simbolo della rottura di ogni possibile *comunicazione* tra gli uomini, dell'*insensatezza* del linguaggio umano “ora fare filosofia – scrive Jaspers nella sua *Autobiografia filosofica* – significa lavorare alla possibilità di una comunicazione universale”⁷⁸, nella consapevolezza che “la verità è ciò che ci unisce”⁷⁹. La comunicazione diventa, dunque, una tra le più feconde poste in gioco filosofiche lasciate in eredità dalla riflessione di Karl Jaspers, che così scrive:

La comunicazione è la sfera del possibile, della libertà. L'illimitata disponibilità per la comunicazione non è la conseguenza di un sapere, ma la decisione verso una strada nell'esser uomo. Il pensiero della comunicazione non è utopia, ma fede. Si pone per ognuno la domanda se egli vi aspira o se vi crede non come a qualcosa posto nell'aldilà, ma come a qualcosa interamente presente: alla possibilità in noi uomini di vivere realmente uno con l'altro, di parlare uno con l'altro, di trovarsi nella verità in virtù di questo vincolo reciproco e di divenire propriamente se stessi lungo questa strada.⁸⁰

L'unità globale del mondo, dal punto di vista economico, politico, finanziario, impone l'esigenza di un nuovo orientamento esistenziale che, senza dimenticare la particolarità dei punti di vista tradizionali nazionali e europeo-occidentali, sia al contempo in grado di oltrepassarli. Alla proposta di una filosofia della storia *universale* fa da controcanto, nel pensiero jaspersiano, l'esigenza di una transizione “dal tramonto della filosofia europea all'aurora della filosofia mondiale”⁸¹.

Delineare i tratti di una *Weltphilosophie* significa per il filosofo tedesco rinvigorire una *Vernunftphilosophie* (filosofia della ragione), espressione storica e concreta di quell'*existenzphilosophie* che da sempre aveva accompagnato la sua riflessione. Essa è testimonianza dell'esigenza di rivitalizzare il lavoro della ragione, “atmosfera” e “respiro”⁸² dell'esistenza,

⁷⁸ K. Jaspers, *Philosophische Autobiographie* in *Karl Jaspers*, hrsg. von P.A. Schilpp, Kohlhammer, Stuttgart 1957, pp. 1-80; tr. it. *Autobiografia filosofica*, Morano, Napoli 1969, p. 148.

⁷⁹ Ivi, p. 155.

⁸⁰ K. Jaspers, *La fede filosofica*, cit., p. 158.

⁸¹ Ivi, p. 120.

⁸² K. Jaspers, *Ist die Philosophie am Ende? Ein Gespräch mit Willy Hochkeppel über die Zukunft der Philosophie* (1961), in *Provokationem. Gespräche und Interviews*, Piper, München 1969, pp. 11-32, p. 29.

nella sua critica contro ogni sistema chiuso, esclusivizzante o definitivo e in una riappropriazione polemica e creativa del passato.

Hans Saner, curatore degli inediti di Jaspers, ha presentato, pur per cenni, l'intero programma di lavoro jaspersiano sulla storia *mondiale* della filosofia, di cui esistono ben 22.000 fogli manoscritti, che si sviluppa lungo alcune direttrici fondamentali⁸³.

Il primo libro (cui appartengono circa tremila fogli) intendeva costituire una storia *cronologica* comprendente una ricostruzione storiografica delle culture originate e sviluppate in Grecia, in India e in Cina, insieme all'analisi della filosofia occidentale dal cristianesimo sino all'età contemporanea. L'appendice era poi riservata a dimostrare i parallelismi e le relazioni tra le distinte epoche storiche e il loro significato nell'elaborazione del pensiero dei singoli filosofi. La seconda indagine (che computa circa 35.000 fogli, arricchiti di note e lunghe citazioni) voleva invece tracciare una storia dei *problemi filosofici* attraverso l'analisi delle grandi questioni sul mondo, su Dio, sull'Essere. La terza indagine, infine, trova una prima sintesi problematica in *Die großen Philosophen*⁸⁴, cui è necessario aggiungere 14.000 cartelle inedite comprendenti monografie di numerosi pensatori. Tema della quarta ricerca sarebbe dovuto essere il rapporto tra la filosofia e il linguaggio, il mito, la religione, la poesia e l'arte. Il volume seguente – suddiviso in 650 cartelle – avrebbe dovuto rispondere all'interrogativo: “come si è realizzata la filosofia nella sua storia”?, rintracciando riflessioni filosofiche anche in opere di natura differente, in particolar modo letteraria, pedagogica e politica. Il sesto libro, infine, sarebbe stato destinato a focalizzare le incomprensioni storiche sorte tra i filosofi, che hanno intessuto la trama delle *querelles* più conosciute e più interessanti per lo sviluppo spirituale dell'uomo⁸⁵.

⁸³ Cfr. H. Saner, *Karl Jaspers*, Rowohlt, Hamburg 1970. Per un approfondimento critico sul tema si veda P. Ricci Sindoni, *Jaspers e l'idea di una filosofia universale: etica e avvenire dell'umanità* in G. Cantillo (a cura di), *Karl Jaspers. Esistenza e Trascendenza*, Pro Civitate Christiana, Assisi 1989, in part. pp. 176-180.

⁸⁴ K. Jaspers, *Die großen Philosophen*, Piper, München 1957; tr. it. *I grandi filosofi*, Longanesi, Milano 1973.

⁸⁵ Si veda quanto commenta Saner: “Nell'ottica cronologica la storia universale della filosofia diventava così una storia universale degli spazi culturali, delle epoche e dei periodi. Dal punto di vista oggettivo diventava una storia universale dei contenuti filosofici, nell'ottica personale diventava una storia dei grandi filosofi, in quella genetica storia universale del fondamento e della ripercussione della filosofia nel linguaggio, nel mito, nella religione e nell'arte, nella prospettiva pragmatica diventava la storia universale nella realizzazione concreta della filosofia; e in quella dinamica la storia del conflitto e della comunicazione. La storia universale della ragione doveva dunque essere scritta sei volte, nella speranza che poi una sorta di “sinossi” di tutte queste storie permettesse di guardare la stessa totalità intrinsecamente non rappresentabile” (H. Saner, *K. Jaspers. Dall'ampiezza della ragione e dell'attendibilità dell'agire*, in E. Nordhofen (a cura di), *Filosofi del 900*, Einaudi, Torino 1988, p. 117).

Tale imponente progetto diventa allora testimonianza di una *filosofia universale a-venire*, innervata nella fede nell'universalità della ragione, intesa come creazione di uno *spazio comune* in cui comunicare con tutti. All'interno di esso, si avverte il *senso* di un'unità profonda che si realizza come "religione invisibile", "una sorta di repubblica metafisica dei sapienti, dove essi s'incontrano tra loro come fanno i filosofi della Scuola d'Atene nell'affresco di Raffaello"⁸⁶. Tale "regno degli spiriti che si incontrano e si appartengono l'un l'altro" è allora alla base di un'ideale "repubblica noumenica"⁸⁷ fondata sulla libertà della comunicazione illimitata tra le esistenze. "Di fronte a una catastrofe mondiale – così si legge in *I grandi filosofi* – che minaccia la stessa esistenza degli uomini, abbiamo oggi la possibilità di assimilare, sulla base del sapere storico, le grandi esperienze del passato e di tradurle in forze presenti"⁸⁸. A partire da tali presupposti, la storia *mondiale* della filosofia dovrebbe favorire, attraverso il retaggio delle grandi personalità decisive, la formazione di una coscienza storico-filosofica condivisa.

Oltre la tesi di un'origine unilaterale – *solo* occidentale – della filosofia, è allora possibile pensare una storia dei *grandi filosofi* (*Die großen Philosophen*) come riflessione sulle personalità decisive che, in Occidente come in Oriente, hanno realizzato quel distanziamento critico dalla realtà e quell'apertura alla Trascendenza di cui si è nutrito il *respiro* dell'età assiale. L'azione millenaria di Socrate, Buddha, Confucio, Zarathustra, i profeti vetero-testamentari, Gesù ha avuto un'efficacia storica e spirituale di ampiezza e profondità incomparabili. Essi sono "eterni contemporanei", poiché, come scrive Jaspers, "il grande non è già colui che coglie il suo tempo in pensieri, ma colui che attraverso ciò lambisce l'eternità"⁸⁹.

L'unicità e insostituibilità del loro essere li caratterizza *oltre* ogni tipo storico o ogni classificazione sociologica: è per questo che la critica storico-filologica, per quanto centrale nel vaglio critico delle fonti, non è in grado di cogliere una realtà che la trascende, la realtà "sovra-storica" e spirituale che li caratterizza, la "potenza straordinaria esercitata dal loro essere sui loro ambienti e all'eco suscitata presso i posteri"⁹⁰. Nella loro unicità si rivela la "storicità universale dell'essere umano"⁹¹ la realizzazione delle sue possibilità più autentiche⁹².

⁸⁶ K. Jaspers, *Origine e senso della storia*, cit., p. 298.

⁸⁷ K. Jaspers, *I grandi filosofi*, cit., p. 621.

⁸⁸ Ivi, p. 120.

⁸⁹ Ivi, p. 117.

⁹⁰ Cfr. K. Jaspers, *Socrate, Buddha, Confucio, Gesù. Le personalità decisive*, Fazi Editore, Roma 2013, pp. 180-184.

⁹¹ Ivi, p. 186.

⁹² La proposta di una storia della filosofia innervata sulle grandi personalità e sugli individui *sovra-storici* generò non poche polemiche e malintesi: in particolar modo durante

Così scrive Jaspers:

Essi non rispondono al *tipo profetico* che si distingue per visioni o estasi [...]. Sono profeti in senso più grande: qualcosa in essi si è lacerato. Il mondo non è in ordine. Si sente e si esige una trasformazione radicale. [...]. Parlano per metafore, per contraddizioni dialettiche, per repliche dialogiche, senza fissare nulla. Indicano quel che si deve fare, ma in modo che non possa esser conseguito mediante una tecnica [...]. Irrompono le consuetudini, le verità comunemente ammesse, il semplice oggetto di pensiero. Creano uno spazio nuovo con possibilità nuove e lo popolano di tentativi che non vengono mai portati a compimento definitivo.⁹³

Pur non essendo tradizionalmente solo *filosofi*, essi possono essere definiti secondo Jaspers degli *iniziatori*, con tutta la complessità semantica che il concetto assume *dopo e attraverso* la filosofia arendtiana: “la loro grandezza – scrive il filosofo – si rivela nel fatto che non cessano mai di agire”⁹⁴ e, in ascolto della loro eredità, sul loro *esempio*, è possibile per tutti gli uomini generare *nuovi inizi*⁹⁵. L'essenziale della loro esistenza è, infatti, essenziale alla filosofia, in quanto costituisce una fonte del filosofare: “ognuno costituisce per noi un questionare che non ci dà pace”⁹⁶, scrive Jaspers. Non si può infatti, non esser colti dalla loro “esigenza di serietà”, nel progresso di pensiero (Socrate), nella meditazione e in un modo di vivere a essa adeguato (Buddha), in un'educazione che è più di

la giornata di Studi Unesco 2005, Amadou Mathar MBow e Frank Blankart assieme ad alcuni rappresentanti dell'America Latina, dell'Africa e della Russia, colsero il tentativo jaspersiano come il progetto aristocratico della cultura dominante di stampo occidentale, ancora protesa nel suo inconsapevole istinto alla colonizzazione culturale. Basti pensare, in tal senso, al riconoscimento di un'età assiale “solo” in Europa, Cina e India e all'esclusione delle “premesse” filosofiche dal pensiero africano o da quello americano. A tali critiche doveroso aggiungere l'analisi di Habermas: il filosofo tedesco contestava l'*ethos* aristocratico di Jaspers, figlio di quell'interiorità del protestantesimo tedesco che aveva generato – come argomenta Habermas – l'astrattezza dei principi, la scarsa incisività degli appelli moralistici, la non conclusività dei progetti politici e, non da ultimo, una “storia della filosofia *senza storia*”, radicata in un'astratta comunicazione degli spiriti in una platonica repubblica metafisica che non tiene in considerazione la centralità dei rapporti culturali, sociali e storici nell'elaborazione filosofica (cfr. J. Habermas, *Die Gestalten der Wahrheit* [1958], in *Karl Jaspers in der Discussion*, hrsg. Von Saner, Piper, München 1973, pp. 309-316). L'indagine di Saner, che inserisce il progetto de *I grandi filosofi* all'interno di un complesso di studi più imponente e rigoroso, permette di rispondere costruttivamente a queste critiche.

⁹³ K Jaspers, *Socrate, Buddha, Confucio, Gesù. Le personalità decisive*, cit., pp. 187-188.

⁹⁴ Ivi, p. 185.

⁹⁵ A tal proposito, non è soltanto una coincidenza che il volume *Die großen Philosophen* sia stato pubblicato nel 1957, appena un anno prima la pubblicazione di *Human condition* (1958) in cui le idee arendtiane di *nascita, inizio e azione* trovarono una sintesi coerente e approfondita.

⁹⁶ K Jaspers, *Socrate, Buddha, Confucio, Gesù. Le personalità decisive*, cit., p. 198.

un semplice apprendimento (Confucio), con l'abbandonarsi alla volontà di Dio senza considerazione del mondo (Gesù)⁹⁷.

Non si può, dunque, non concordare con Arendt quando, alla notizia della prima stampa del corso di filosofia della storia tenuto da Jaspers nel 1948 esclama: "Certo, Lei fa davvero di nuovo venire voglia di essere cittadini del mondo o, più esattamente, lo rende nuovamente possibile"⁹⁸. È l'autentico spirito repubblicano e democratico a costituire, infatti il più grande lascito della riflessione filosofica e politica jaspersiana, quel "legame tra libertà, ragione e comunicazione che Karl Jaspers ha esemplarmente rappresentato con la propria vita, per poi descriverlo con gli strumenti della riflessione"⁹⁹. Le proposte di una nuova *filosofia* della storia e, in contro canto, di una nuova *storia* della filosofia, pur rimasti nella loro unitarietà di intenti un progetto inconcluso, sono testimonianza dell'impegno etico e politico dello studioso tedesco verso la realizzazione di un ordine internazionale pacifico e di una cittadinanza cosmopolitica, condizione di possibilità fondamentali per quella *seconda età assiale* il cui anelito spira nella nostra attualità. Essa diventa dunque la posta in gioco degli ultimi scritti jaspersiani che, in quello che ormai possiamo definire un "lessico comune" con l'allieva Arendt, sono animati dalla condivisa fiducia che ogni uomo possa essere *nuovo inizio*¹⁰⁰.

⁹⁷ Cfr. *ivi*, p. 188.

⁹⁸ H. Arendt, K. Jaspers, *Carteggio 1926-1969*, cit., p. 234.

⁹⁹ *Ivi*, p. 238.

¹⁰⁰ Cfr. H. Arendt, *Vita activa*, cit., pp. 263-264.

Lars Rensmann*

Antinomies of Progress.

Notes on Adorno's Critical Theory and the Concept of Progress in Our Time

Abstract

Today, the idea of progress is confronted with four distinct but interrelated challenges: global authoritarian regressions and democratic backsliding; the looming climate catastrophe; the partial dissolution of recognizable boundaries between human agency and artificial intelligence; and omnipresent forms of digital reproduction through which seemingly every human interaction is identified, measured, counted, objectified, and valorized. Turning especially to Adorno's essay on "Progress," I address these four steep challenges facing contemporary society and their meaning for (moral) progress from the perspective of Frankfurt School Critical Theory. Rather than resigning to a negative telos and the realistic threat of disaster, I propose a critical and dialectical notion of (both moral and social) progress that takes inspiration from Adorno's work. It contains and upholds the possibility of genuine betterment, of averting catastrophe, and of redemption while critically reflecting on the conditions and societal trends towards destruction and what I call hyperreification. I argue that Adorno's idea of progress continues to shed light on problems and antinomies in the contemporary age of unreason, recognizing both its entanglement in society and its critical qualities.

Keywords

Artificial Intelligence; Humanity; Moral Progress; Posthumanism; Social Media

"As little as humanity *tel quel* progresses by the advertising slogans of the ever new and improved, so little can there be an idea of progress without the idea of humanity."
Theodor W. Adorno, *Progress* (1964)

1. Introduction: On Answering the Question "What is Progress?"

In his famous essay on progress, initially delivered as a lecture at the Philosophers' Congress in Münster in 1962, Theodor W. Adorno laid

* University of Passau.



out the contours of a dialectical notion of progress in “the age of catastrophe,” as he aptly put it (Adorno 1998a, p. 147). In so doing, Adorno closely linked the idea of progress, on the one hand, to several key concepts indebted to Kantian moral and political philosophy, featuring reason, enlightenment and, most importantly, humanity, as well as the theological image of redemption, to which, according to Adorno, the idea of progress ultimately points. For him, genuine progress was needed, as “no good can exist” without progress (Adorno 1998a, p. 147). As there cannot be any good, no morality without progress, progress without morality, and the idea of the good or of human betterment, seems equally impossible and certainly undesirable. Against the backdrop of a society which is, for Critical Theory, constitutively shaped by social domination, the good life is intimately linked to progress, while moral progress appears to be a tautology of sorts.

Yet, on the other hand, Adorno also stressed tensions of “progress” both as a concept and as an empirical reality: its profound antinomies in our empirical world, which is full of ambivalences and contradictions, and its entanglement in a global society actually facing humanity’s extinction—a society which celebrates progress while progressing towards disaster. Without throwing the baby out with the bathwater and giving up on the promise of progress altogether, Adorno’s Critical Theory hereby criticizes a seemingly omnipresent “fetishization of progress” (Adorno 1998a, p. 160) in the modern world of permanent reifications and transformations—a “progress” that is restricted to refined techniques, technological advances, commodity value and valorization, as well as increased mastery over nature and perfected social domination, yet sold to us citizens as true progress altogether. Adorno thus rejects a ubiquitously advertised progress, which is deceptive and which furthers, actually, humanity’s delusions about progress while blindly reproducing conformist subjugations to social pressure, comprehensive adaptations to societal demands, and ultimately humanity’s quite possible extinction. Contemporary apologists of blind progress of our time, including so-called posthumanists, just as those during Adorno’s, tend to render resistance to such progress “conservative” and “backward-looking”, fallen out of time. However, in what seems to have fallen out of time, a Critical Theory of progress indebted to Adorno, by contrast, recognizes the critical potential of progress in conditions, ideas, claims and subjectivity that societal and technological “progress” seemingly tends to undermine: human freedom, democracy, subjectivity, cosmopolitan solidarity, and humanity.

In our contemporary world, how can we answer the question: “What is progress?” How can we uphold a critical notion of progress today in view of multiple challenges to the very conditions of (moral and social) progress, and what may we learn from Adorno’s reflections in order to

advance such a concept? In what follows, I will reread some of Adorno's reflections on the notion of progress, as developed in his influential lecture, and shine light on his dialectical understanding of the concept in view of our contemporary condition—of what I have called a new age of unreason (Rensmann 2017a). Our present time is shaped by authoritarian regressions and democratic backsliding the world over; an autocratization trend bolstering authoritarianism in both democracies and autocracies that is in various ways reminiscent of the rise of the fascist movements Adorno witnessed in the 1930s, before the catastrophe of the Shoah. Adorno's arguments about progress will also be situated in the contemporary context of the climate crisis, which brings our planet ever closer to extinction; and be discussed in view of the enormous technological transformations we are facing, featuring the digitization of our communication and socialization as well as the advent of ubiquitous artificial intelligence. Any meaningful reconstruction and actualization of Critical Theory's notion of progress needs to take account of these changing conditions. However, I will argue that Adorno's dialectical understanding, while facing limitations, offers important paths for preparing a critical notion of progress today, illuminating some persistent antinomies and problems in a different, new, but no less contradictory "age of both utopian and absolutely destructive possibilities" (Adorno 1998a, p. 143).

First, I will turn to Adorno's understanding of the "antinomies of progress", and how these antinomies may play out in our time. I will specifically address and complicate Adorno's dialectical view of progress in an age of digitization, in which subjects are socialized through societal institutions such as TikTok, and in the face of new authoritarianism. Second, I will spell out the significance of humanity, as an ideology and as a claim yet to be universally realized, for any critical notion of progress in light of Adorno's writing and in view of the antihumanistic, relativistic political and intellectual environment such understanding faces today. Against the backdrop of his emphatic cosmopolitan understanding of humanity as an idea still to be realized, I renew the link from Adorno's notion of progress to his call for a new global subject as the necessary condition to avert disaster in the horizon of the climate crisis. Third, I will briefly address the challenging effects of artificial intelligence and robotization and their potential meaning for human subjectivity, humanity, and the notion of moral progress in an Adornian take. And fourth, I will conclude my reflections against this background with a plea against posthumanism in all its technological cyborg and non-technological jellyfish variants. They both ultimately render the inclusive idea of humanity superfluous but without which, I argue, no emphatic idea of progress can exist. Hence, I put forth a plea against post- and metahumanist irrationalism, which constitutes the philosophical soundtrack to an authoritarian age of unreason unfolding in front of our eyes.

2. From the Slingshot to TikTok: The Antinomies of Progress in the Age of Digital Reproduction

For Adorno, the concept of progress is intertwined with the modern condition and its inherently transformative dynamic, both normatively and in terms of societal practices. Claims to progress are part and parcel of the modern world. For a critical theoretical account of progress, then, “it is necessary to scrutinize the category so closely that it loses its semblance and obviousness, both in its positive and negative usage.” However, by the same token, insisting on exactitude only misguides us because of its inherent tensions and ambiguity: “where the impossibility of the unambiguous appertains to the subject matter itself, dogmatic epistemology misses its object, sabotages insight and helps to perpetuate the bad by zealously forbidding reflection upon what ... the consciousness of those entangled would like to discover: whether there is progress.” (Adorno 1998a, p. 143)

For preparing a dialectical notion of progress inspired by Critical Theory and providing an answer to the question of whether there is progress in our time, Adorno offers us some direction. The “antinomian character of progress” (Adorno 1998a, p. 147) today points to the contradictions and ambiguities in the contemporary world but also to the need to avoid writing histories of success or decline. While progress proceeds, on the one hand, from the slingshot to the megabomb, there is also, undeniably, actual progress, on the other hand, “from slavery to the formal freedom of subjects” and “from deprivation to provisions against epidemics and famine” (Adorno 1998a, 148). Most importantly, it is essential for a Critical Theory of progress, *pace* Adorno, to not renounce thinking the idea of progress altogether, even in the “age of catastrophe”: “Progress should be no more ontologized, unreflectively ascribed to Being, than should decline, though indeed the latter seems to be the preference of recent philosophy. Too little of what is good has the power in the world for progress to be expressed in a predicative judgment about the world, but *there can be no good, not a trace of it, without progress.*” (Adorno 1998a, p. 147, emphasis added)

The idea of progress, as a promise yet to be realized, has both a modest and a redemptive quality in Adorno’s take. On the one hand, it means averting total disaster, catastrophe, “radical evil” (Adorno 1998a, p. 160), maybe allowing, at last, for breathing a “sigh of relief” (Adorno 1998a, p. 144). On the other hand, progress for Adorno is unthinkable without the redemptive idea of the emancipation of humanity and the end of societal domination. Progress allows the “possibility of redemption to flash up” (Adorno 1998a, p. 148) and thus contradicts what is. And such progress requires self-reflective reason that can possibly transition to praxis: “The

explosive tendency of progress is not merely the Other to the movement of a progressing domination of nature, not just its abstract negation; rather it requires the unfolding of reason through the very domination of nature. Only reason, the principle of societal domination inverted into the subject, would be capable of abolishing this domination.” (Adorno 1998a, p. 152)

Progress, says Adorno, is entangled in reason, from which it cannot be separated if progress is to have any meaning at all. Progress is thus also entangled in society and history, and it does not exist outside either. Without society, “the notion of progress would be completely empty; all its elements are abstracted from society.” (Adorno 1998a, p. 148) For Adorno’s Critical Theory societal, political, and historical dimensions of progress can never be neatly distinguished from the moral and ethical ones, which are the primary concern of moral philosophy. The latter is embedded in and complicated by the social and historical conditions under which it operates. It is the neglect of this insight what is most problematic, in Adorno’s take, about Kant’s formalism and moral philosophy—and his notion of moral progress.

Hence, the concept of progress, according to Adorno, “is philosophical in that it articulates the movement of society while at the same time contradicting it.” (Adorno 1998a, p. 148) Progress belongs to the quasi-mythological circulation of modern capitalist society and the dynamic of modern social domination; it is part of an “antagonistic unity of movement and standstill” (Adorno 1998a, p. 160), the “interlocking of ever-same and the new in the exchange relation” which “manifests itself in the imagines of progress under bourgeois industrialism.” (Adorno 1998a, p. 159) Yet, progress also has the potential to explode the mythical spell to which it belongs: “Progress means: to step out of the magic spell, even out of the spell of progress that is itself nature, in that humanity becomes aware of its own inbred nature and brings to a halt the domination it exacts upon nature and through which domination by nature continues. In this way it could be said that progress occurs where it ends.” (Adorno 1998a, p. 150)

“Part of the dialectic of progress,” argues Adorno, “is that historical setbacks, which themselves are instigated by the principle of progress...also provide the condition needed for humanity to avert them in the future.” (Adorno 1998a, p. 154) Progress can, indeed, be conceived as the “resistance to the perpetual danger of relapse.” (Adorno 1998a, p. 160) But the dangers of universal regression, of authoritarian backsliding and even of human/environmental catastrophe are real, and today, as at many critical junctures in the past, (technological and organizational) progress actually seems to be allied with such political and societal regression.

In contemporary digital social interactions, which are partly governed by and embedded in social media corporations' algorithms, and which are constantly, "progressively" perfected, enhanced, or directly generated by artificial intelligence, every image seems stylized, staged, and manufactured for the purpose of social or economic valorization. Moreover, technological progress here also means that every reaction is identified, objectified, measured, counted, and valorized for economic gain. This is digital "progress" for you. Rather than rendering objectification, reification, and utilization obsolete as powerful social mechanisms to which all members of modern societies are subjugated, social life under conditions of 21st century global digital capitalism has become increasingly commodified or, as I call it, *hyperreified*. Initial studies indicate that the digital restructuring of the public sphere tends to destroy a vivid public realm and replace it with a seemingly never-ending flood of disinformation, post-factualism, and tribalism.

Moreover, current social media, a result of technological progress serving capitalist ends, now constitutes a key institution of socialization. This has also dramatically destructive effects on the cognitive and moral capacities of citizens and denizens as consumers, and particularly children. Consuming TikTok videos for hours, as many children and young adults do, literally kills your brain. Alongside progressing the loss of key critical faculties, consciousness and conscience, I suggest, that such hyperreification tends to engender and reward both social conformism and authoritarian aggressions. By and large, empirical research suggests that the technological progress of digitization further weakens the competences of weakened human modern subjects to resist stupidity, moral bankruptcy and, most importantly, blind authoritarian submission towards both general societal pressures and towards dictators or wanna-be autocrats among the "technologically educated masses" (Horkheimer & Adorno 2002, p. xvi). The result may be increasingly weakened, authoritarian personalities in present and future generations: amorphous, totally flexible, stupefied. The contemporary, hyperreified structure of, and interactions on, social media hence appear to nurture new forms of authoritarianism among younger generational cohorts. They are doing so by direct, unmediated technological means, rather mediated forms of primary and secondary socialization. Rather than democratizing socialization, education, and the public, the digital transformation of the social and public sphere, dominated by powerful corporations and their algorithms, has both autocratized and disinhibited social interactions. It points to an increasingly radical expropriation of subjectivity and weakening of critical faculties or competences, including moral conscience and consciousness—far beyond what Adorno could have imagined.

This fits neatly the societal demands and control mechanisms of neo-liberal global capitalism. It subjugates humans globally to the requirements of global production and trade in the name of progress (and freedom, for that matter). Under transformed societal and digitized conditions, individuals have seemingly become ever more ready to conform to all-pervasive new economic demands for flexibility, voluntary self-exploitation, and constant self-enhancement (Jäger 2022, Bueno 2021; Gandesha 2018). Adorno's 'radio generation' of the 1940s, then, has arguably been replaced by the 'social media generation(s)' of our time. What may have been exaggerated claims in Adorno's time, namely that socialization processes have largely expropriated the nuclear family and its mediations while society directly reaches into all aspects of socialization from early childhood onwards, seems to have come to fruition in the age of digital technology and social media. Induced by the new stage of technological progress for profit, technology-driven social and communicative conditions of modern digital capitalism shape an individual's subjectivity, composition, and reactions on the earliest and most intimate levels. Such conditions may engender the blind acceptance of both authoritarian political systems and a social world without thought, driven by the abstract authority of powerful and omnipresent technological companies. This development signifies the progress from the slingshot to TikTok, the destructive possibilities of which are only foreshadowed so far, though the writing is on the wall. It points to a situation of "universal regression that allies itself with progress today." (Adorno 1998a, p. 160)

Still, it would be an uncritical and undialectical notion of progress to view technological progress with all its destructive possibilities solely in terms of a history of decline. While seemingly outdated social practices that enable critical thinking, such as reading books, may be more beneficial to preserving the condition of possibility for progress than consuming social media, technological advances of the digital age may also be put to productive use once they are no longer under the spell of valorization and the exchange principle.

3. Averting Total Disaster: Moral Progress in the Face of Humanity's Extinction and the Need for Cosmopolitan Solidarity, Democracy, and Critique

Answering the question whether genuine progress exists today means ultimately no less than answering the question "whether humanity is capable of preventing catastrophe." (Adorno 1998a, p. 144). Averting total disaster as the modest measure of progress: This was true for Adorno,

and it is no less evident in an age of climate crisis, of which Adorno could not have had any idea.

Possibly even more self-evident than during his time is the need for an intervening global subject against the globalized condition of environmentally destructive capitalism undermining the future of humankind and life on the planet. Adorno calls for the rehabilitation of the “Kantian universal cosmopolitan concept” (Adorno 1998a, p. 145) in the face of global disaster, rather than the return to authoritarian-nationalist particularism, if progress as averting catastrophe stands any chance of coming to fruition (Rensmann 2016): “The forms of humanity’s own global societal constitution threaten its life. The possibility of progress, of averting the most extreme, total disaster, has migrated to the global subject alone.” (Adorno 1998a, p. 144) For Adorno, the avoidance of global catastrophe thus requires the free and “rational establishment of the whole society as humanity” (Adorno 1998a, p. 144), and the realization of Kant’s cosmopolitanism.¹

Progress, then, “would be the very establishment of humanity in the first place, whose prospect opens up in the face of its extinction.” (Adorno 1998a, p. 145) Progress understood as the rational establishment of society as humanity, that is, a global subject capable of preventing disaster of a global scope, is inevitably attached to the cosmopolitan idea of humanity and cosmopolitan solidarity (Basnett 2023; Rensmann 2016). Through Adorno’s lens, as indicated, there cannot be an idea of progress “without the idea of humanity” (Adorno 1998a, p. 145) in all of its meanings and connotations; indeed, genuine progress, from a Critical Theory perspective, can only be conceived as progress towards—yet unrealized—humanity. This also requires, of course, critical reflection on the conditions that have caused humanity’s crisis and brought humanity to the brink of total destruction, once again.

Cosmopolitan solidarity and the realization of humanity today undoubtedly depend on ensuring the conditions for humanity’s reproduction: to do everything possible to save the planet, the environment, and avert the climate catastrophe. This ultimately requires a different, self-reflective rather than entirely objectifying relationship of humanity to (both inner and external) nature, instead of the blind “triumph of society over nature—a triumph which transforms everything into mere nature.” (Horkheimer & Adorno 2002, p. 153) This means that there is the need

¹ Of course, within Kantian enlightenment, which “first of all puts progress towards humanity in people’s own hands and thereby concretizes the idea of progress as one to be realized, lurks the conformist confirmation of what merely exists.” Yet critically recognizing Kant’s limitations, including the subtle “conformist confirmation” of what exists within Kant’s philosophy, does not disqualify Kant’s cosmopolitan idea(l).

to intervene into the dialectic of blind, objectifying mastery of humankind over nature that deprives human subjects of their own critical potential and subjectivity and turns everything equally into jellyfish without critical consciousness and emancipatory agency, including the agency to prevent humanity's extinction. Climate change, which Adorno could not anticipate but which is one of the consequences of the blind mastery, reification, and exploitative relationship of society to the natural world, makes the need for progress towards humanity acting as a self-conscious global subject, the need for the rational establishment of the whole society as humanity, as Adorno puts it, all the more pressing.

The establishment of humanity as a global subject, which for Adorno is a condition for moral progress and for the possibility to avert disaster, also implies a commitment to democracy as the framework for critique. Democracy provides the condition—the space—for critical and free public reflection and guarantees individual and collective rights. Without democracy—“nothing less than defined by critique”—and without the “conception of the separation of powers” (Adorno 1998b, p. 281), there can also be no moral philosophy in any meaningful way that points to actual (moral) progress in society. Without the context of democracy and its prerequisite, “political maturity” (Adorno 1998b, p. 281), then, there can also be little chance of averting disaster—the genuine form of progress one can realistically hope for in our time.

An abstract moral philosophy of progress in our time has become even more dubious, from a Critical Theory perspective, than it was in the past. A moral philosophy of progress can no longer avoid reflecting on the need for a new cosmopolitan solidarity that establishes humanity, i.e. a democratic, free and just global society, and extending solidarity to animals and the natural world we inhabit. Likewise, Adorno's Critical Theory presses us to recognize that any moral philosophy of progress risks becoming ideological and delusional—and arguably more so today than in Adorno's time—if it does not reflect on the actual political and societal contexts and material conditions under which it operates. Indeed, today neither progress, nor moral philosophizing about progress that does not fully betray the very idea of progress as linked to humanity, as an ideal and practice *pace* Kant, can happen if, for instance, grave injustices and autocratic conditions cannot be addressed, confronted, and criticized. Indeed, in the executioner's house, moral philosophers today need to talk about the rope—and of societal conditions further advancing destructive dynamics of progress which tend to eliminate the conditions of possibility for genuine progress, for human emancipation and humanity to emerge.

How do you practice moral philosophy, for instance, in China today, which has in recent years eclipsed all hitherto remaining spaces, in

universities and the public realm at large, to prevent any critique of the Chinese regime? The personalist dictatorship of Xi Jinping regime silences and jails dissidents (and worse), including those who display any disagreement within the Communist ruling elite; the regime terrorizes its population, radically curtails free speech, persecutes the Uighurs and other minorities, and has built one of the most authoritarian, oppressive, immoral systems of rule in the 21st century. Under Xi, a system of near total surveillance has been established. It is enhanced by digital means designed to extinguish the very idea of human freedom and any hope for betterment, let alone genuine human progress; a system which can be called, *pace* Shoshana Zuboff's concept of digital "surveillance capitalism" (2019), digital *surveillance dictatorship*. Is it possible to actually advance moral philosophy in this context, and in particular applied moral philosophy? How do you teach ethics of Artificial Intelligence (AI) in China, where the ruling regime uses AI for its domestically enforced and globally oriented "tech-enhanced authoritarianism" (Hoffman 2022, p. 76) and its so-called "social credit system" (see Yang 2019), which detects, registers, and sanctions every human behavior based on total surveillance employing face recognition systems? The European Union, for instance, just established ethical guidelines on AI based on the advice, among others, of moral philosophers. Most prominently, this features a complete ban on AI-based face recognition within the EU: "The new rules ban certain AI applications that threaten citizens' rights, including biometric categorization systems based on sensitive characteristics and untargeted scraping of facial images from the internet or CCTV footage to create facial recognition databases. Emotion recognition in the workplace and schools, social scoring, predictive policing (when it is based solely on profiling a person or assessing their characteristics), and AI that manipulates human behavior or exploits people's vulnerabilities will also be forbidden." (European Parliament 2024) For moral and ethical reasons, hence, the EU bans AI uses which are central to the Chinese regime's totalitarian project.

So how can we engage with Chinese scholars and universities today in any meaningful way that allows for open debate and progress, when the most basic forms of critique, ethical reflection, and moral philosophy are impossible and lead to political persecution? This constitutes an empirical and moral dilemma which needs to be the subject of reflection. The dilemma does not imply advocating for boycotting Chinese scholars, or principled disengagement with those raising objections to torture and persecution, however veiled. Yet the idea that exchange with authoritarian or totalitarian institutions and their representatives is always beneficial and leads to progress and opens up paths to humanity and societal betterment, no matter the parameters that these institutions establish, is

empirically delusional and in moral terms deeply problematic. Simply accepting radically oppressive or totalitarian conditions, the darkness at noon around you, is not a sign of cultural openness, diversity, tolerance, or postcolonial inclusivity—no matter if the regime is European, Asian, or African. To the contrary, conforming to such conditions in the name of cultural difference, or pretending they do not exist and do not affect moral philosophy, constitutes a profound moral failure. It means a general betrayal of a Kantian cosmopolitanism that extends to every human being—and should today reach beyond humans—as well as a particular betrayal of those human beings and groups who are arrested, persecuted, deprived of their human rights, and struggle for their survival, if in China, Iran, Russia, or elsewhere.

In this spirit, Adorno also sides with Kantian formalism because it recognizes “the bourgeois equality of all subjects” that can be contrasted to “the allegedly a priori differences that are supposed to exist between people according to fascist principles.” (Adorno 2006, p. 252-3) In the “universal legal norm...despite and because of its abstractness, there survives...something of substance: the egalitarian idea.” (Adorno 1973, p. 236) A difference-sensitive cosmopolitanism indebted to or inspired by Adorno can never be oblivious to the violation of basic human rights and dignity, the substantive egalitarian idea that is the condition of being different without fear (Rensmann 2017b). Universal human rights remain true as an idea, hitherto unfulfilled. The critical image of human rights matters even if it still falls short in bringing about justice and freedom; the reification of alleged cultural differences and collective particularisms that trump such universal rights, by contrast, violates the very idea of humanity, freedom, and progress by locking humans into identities from which they allegedly cannot escape. Without humanity as an idea and practice and without global human responsibility, it is difficult to even imagine what progress could mean in our time. This is what Adorno already recognized well before the current age of authoritarian unreason and the looming environmental disaster.

4. Of Humans, Robots, and Artificial Intelligence: Progress, Autonomy, and the Dissolution of the Boundaries of the (Moral) Self

Just as Adorno recognizes all sorts of actual societal progress, such as the progress towards democratic constitutions and separations of power, which enable the formal freedom of subjects as well as public freedom and critique, Adorno’s dialectical notion of progress also emphasizes, as mentioned, various forms of social progress in many parts of the world, including measures against “epidemics and famine”. His conception of progress

is thus not simply entangled in a *Verfallsgeschichte*, a history of decline. It would be equally misguided to construe Adorno as an opponent to technological and scientific progress. Taking inspiration from Adorno's notion of progress, a Critical Theory of progress I gesture towards recognizes enormous progress especially in the medical field. Modern society is capable of developing crucial vaccines against new global pandemics in a stunningly short amount of time, for instance. All sorts of cancer and other diseases can be cured today, raising life expectancy in many parts of the world. And the advancement of robotic and other medical technologies, including AI, can help humans better detect diseases, conduct better surgeries, and live better lives, even after grave injuries. Moral progress, in this way, can be engendered by technological progress if it makes human life better (and animal and planetary life, for that matter). In general, Critical Theory recognizes that medical technology in the service of human subjects (or animals), humanity, and the good, can mean progress, even if it was originally generated in contexts of social domination and production for profit. The primary problem here is not the progress of medical technologies and science helping humans live a better life but, as with other social goods, their fair and equal global distribution and accessibility (or the actual lack thereof under conditions of neo-liberal global capitalism).

Technological progress today, featuring AI-based technologies, certainly challenges the boundaries of the moral subjects as well as human agency and autonomy as the condition to act morally, on a new level. The challenge is also epistemological, as it becomes increasingly difficult, so it seems, to draw strict boundaries between human morality, agency, and autonomy, on the one hand, and artificial activity on the other (including the "independent" capacity of chatboxes to write better papers than students). Alongside conventional limits of the self, the emergence and increasing relevance of humanoid robotics, nanorobotics, and other new technologies thus also seemingly dissolve conventional human ethics and epistemologies. In particular, conceptions of moral and epistemological autonomy are undermined by automatization processes and developments penetrating the boundaries of alleged autonomous subjects and moral selves.

In contrast to the clear-cut boundaries which classical, Kantian or Rawlsian moral philosophy thought to establish and uphold, including strict epistemological ones, however, Adorno's Critical Theory prepares us for the new challenge presented by automatization, robotization, and artificial intelligence. Adorno's work does so considerably better than conventional moral philosophy because Adorno thoroughly illuminates the limits of any conception of individual or moral autonomy, situating the self in its relationship to nature and society and deciphering the claim to autonomy as both true and false.

From the perspective of Critical Theory, autonomy is true as humanity's and the subjects' aspiration to be free from coercion and blind, heteronomous dependency vis-à-vis powerful societal mechanisms—and from being subjected to social domination, including social domination by means of technology (for which also AI can be and is often used today). The idea of autonomy is false and fictitious, however, inasmuch as it suggests an already realized (moral) freedom that can be isolated from its societal conditions and contradictions; and that autonomy can be exercised in disregard of an object by an autonomous subject making independent decisions and judgements, as exemplified by Kantian moral philosophy: What “is decisive in the ego,” Adorno argues in *Negative Dialectics*, “its independence and autonomy, can be judged only in relation to its otherness, to the nonego. Whether or not there is autonomy depends upon its adversary and antithesis, on the object which either grants or denies autonomy to the subject. Detached from an object, autonomy is fictitious.” (Adorno 1973, p. 223) The idea of autonomy, Adorno recognizes, is also fictitious insofar as it replicates constitutive, but false philosophical and epistemological dichotomies between subject and object, identity and non-identity, reason and nature, reason and emotion, consciousness and body, ego and nonego, the moral self and society. We may include: human capacities and technological resources. Adorno writes: “The subjects are not only fused with their own physical nature; a consistent legality holds sway also in the psychological realm, which reflection has laboriously divided from the world of bodies.” (Adorno 1973, p. 221) As Katariina Holma and Hanna-Maija Huhtala show, Adorno's critique of the Kantian notion of (moral) autonomy “attends to the obstacles to genuine autonomy that arise from real-life circumstances.” (Holma & Huhtala 2016, p. 373) These circumstances and conditions change. For Adorno, moral autonomy is always embedded, historical, and relational; it is social and linked to nature, needs, feelings, as well as the objective empirical world—all of this, as well as the evolution of modern technologies, constantly challenges the boundaries of the autonomous self that moral philosophy asserts.

Adorno's dialectical notion of progress rejects an idealist reification of the subject blinded towards its objective natural and societal conditions, and the historical and physical transformations thereof. Reflection makes “the subjects aware of the bounds of their freedom is that they are part of nature, and finally, that they are powerless against society, which has become independent of them” (Adorno 1973, p. 221) Such self-reflection on the dialectically interwoven relationship between the subject's inner nature and its relationship to external nature and society, the awareness of the dialectics of identity and non-identity, equips Critical Theory with an epistemological advantage. This also the case when coping with new

technological developments that profoundly affect conventional boundaries of the human body as well as concepts of moral autonomy and progress.

Facing questions of AI and medical robotics, we can learn from Adorno that the boundaries of the self, moral and bodily, are necessarily fuzzy and complex. They are entangled with changing natural, historical and societal conditions—and thus also intertwined with empirical and moral contradictions. Autonomy, however desirable as emancipation from heteronomy and material constraints, is partly fictitious, limited, constantly shaped by nature and penetrated by the object world; autonomy can certainly never be absolute. The human subject cannot assert hardened boundaries of identity detached from nature or society without deluding itself. In this framework, the progress of new technological conditions affecting human identities and bodies does not present a profoundly new moral or epistemological challenge. Yet, Critical Theory's dialectical framework still insists on the realization of reason, humanity, autonomy, and emancipation as key points of reference for genuine (moral) progress, rather than affirming their dissolution. This also applies to assessing far-reaching technological changes affecting human subjectivity. In fact, insofar as the dissolution of the subject and its integrity serves expanded social control, hyperreification, domination, heteronomy, and instrumental valorization and tears down all differences between subject and object, instead of problematizing a fictitious antagonism between them, such dissolution further undermines the condition of possibility of freedom and (moral) autonomy, i.e. genuine progress.

5. Against Posthumanism: Progress as Regression in a World of Cyborgs and Jellyfish

This leads me to a final, related argument about progress today to which Adorno's Critical Theory may provide a useful approach. It concerns the new trend of posthumanist thought and its relationship to progress. Critical Theory may help shine light on posthumanism, the fashionable philosophical ideology that tends to abstractly affirm and blindly endorse, without regrets, the aforementioned idea of the dissolution of humanity, subjectivity and autonomy, as well as the full-fledged dissolution of ethical and conceptual boundaries between humans and technology, human activity and artificial intelligence, or humans, non-human life and material substrates.

Posthumanism intends to decenter and devalue the human subject. There is some validity to that. Critical Theory also challenges the Promethean human subject which seeks blind mastery over nature and the

inner and external object world, no matter the price subject and nature have to pay for this (Hofstätter 2019). Similarly, some proponents of posthumanism mainly call attention to non-human life and natural conditions on earth (and their ethical meaning) and confront “senseless and self-destructive” (Horkheimer & Adorno 2002, p. 19) forms of domination over nature that are practiced by human subjects. Global human society is indeed on course towards ecological disaster, and would do well to treat animals and other non-human life in a profoundly different way. Progress, understood as economic progress towards growth, can also be rightly questioned and criticized in this light, and there are moral issues that should legitimately be extended beyond the realm of the human species and include animals and the environment, which provides the conditions for life on earth.

Yet, all too often, posthumanism tends to throw out the baby with the bathwater, falling for anti-humanist impulses and rendering humanity superfluous if not justifiably ready for extinction. Such thought represents a form of abstract negation, driven by anti-enlightenment or counter-enlightenment impulses, in its well-known modernist and esoteric variations. Following destructive paths in the history of ideas, they either blindly endorse or reject technological “progress” but equally giving little value to humanity. Erasing the differences between humans, robots, plants, and worms, posthumanism generally denigrates humanity, treating humanity as just one of many natural species or as an outdated concept.

We can distinguish between technological and non-technological, esoteric posthumanism. The first, technological branch of posthumanism, tends to fetishize technological or technical progress. This branch finds expression, for instance, in the “transhuman” endorsement of genetic selection, gene editing and “genetic engineering”, which Stefan Sorgner elevates to humanity’s “most important scientific invention” (Sorgner 2021, p. 61)—however radically such selective editing may march towards inhumanity. Representatives of this branch usually welcome any “posthumanization” of society. This means they seem to applaud human dissolution, by means of robots or artificial intelligence, and seek to grant new robotic entities full membership.

Rather than celebrating this type of “technological progress”, the second, non-technological or esoteric branch of posthumanism fetishizes the beauty of the life of stones and jellyfish, giving up on the notion of progress altogether. Here plants or even bricks tend to be magically equipped with subjectivity, which appears as either equal or superior to human subjectivity. Some ‘new materialist’ posthumanists like Rosi Braidotti at least recognize that some notion of human betterment or advancement, inevitably attached to the concept of humanity and hu-

manism, needs to be upheld, and thus a commitment to human needs, social justice, and human 'becoming' (see Braidotti 2013, p. 29; Bennett 2016, p. 61). Yet even the more reflective representatives of this branch also propose new forms of integration into some nebulous forms of social or natural organisms mimicking nature, which always smacks of justifications for, and has traces of, direct oppression (Hofstätter 2019, p. 8). In this group of scholars, we also find posthumanist new materialists such as Donna Haraway, who proposes a conception of 'compost society'. She endorses "a fuller consideration of nonhuman agency at a material level" that celebrates the total breakdown of boundaries between subject and organism, human and non-human (see Timeto 2021). Such esoteric engagement with stones and jellyfish as partners for dialogue represents a mystical re-enchantment of the world. As Antonia Hofstätter observes, for Haraway a critical thinker like Adorno has thus "little to offer as his 'resolute secularism'... leaves him unable to 'really listen to the squids, the bacteria, and angry old women of Terra/Gaia.'" (Hofstätter 2019, p. 8, quoting Haraway 2016, p. 73-74, endnote 50) Likewise, Francesca Ferrando advances esoteric posthumanism by going fully against the enlightenment while openly replacing moral reason and reflection with "spirituality". She suggests that "the notion of spirituality dramatically broadens our understanding of the posthuman" and claims that "existence, in a spiritual sense, contemplates a non-separation between the inner and outer worlds." (Ferrando 2016, p. 243) The posthumanist "new materialist option," then, ultimately points not just to esoteric spiritual subjectivities projected even to a reenchanting material world but also to the self-destructive, entirely irrational liberation of the environment from the domination of the human species (Hofstätter 2019, p. 8). Such posthumanist irrationalism, passionately directed against anthropocentric thinking and practice, points to philosophical and political forms of "voluntary human extinction" in a posthuman future, conceived as a future entirely without humans. It is a world in which no moral progress is thinkable and possible.

Critical Theory helps illuminate that both forms of posthumanism, call it the cyborg variant and the jellyfish adoration variant, are two sides of the same coin: they render subjectivity, humanity, or reason altogether superfluous, while indulging in witnessing their downfall. Pretending to be either coolly rational by glorifying the alleged superiority of machines over humans, or happily celebrating life forms devoid of all reason in the first place, posthumanism, then, represents the perfect philosophical drumbeat of our time. It ultimately offers an ideology, rather than a philosophical horizon, for an authoritarian new age of unreason. Hand in hand with other resurging authoritarian-irrationalist ideologies and philosophies of our time in this "regressive moment", posthumanism

signifies a contemporary branch of “the mysterious willingness of the technologically educated masses to fall under the spell of any despotism” (Horkheimer & Adorno 2002, p. xvi).

Posthumanism, in all its variations, seems to prepare humans to unreflectively accept their fate, if not their annihilation. Posthumanism, then, seems to either aim at making humans blindly accept whatever technologically supported digital dictatorship enhanced by artificial intelligence may have in store for them—or posthumanism openly advocates for human self-destruction by dismissing all humanity and enlightenment as “oppressive”. In the guise of technically engineered progress, or by advocating the complete negation of the very ideas of progress, reason, and humanity, posthumanism ultimately advances authoritarian regression and submission. They do so in the name of decentering the world by degrading human subjectivity. Contrary to all too common misunderstandings, however, such proposed decentering by denigrating humanity represents the opposite of Critical Theory’s critique of the Promethean subject seeking total mastery over internal and external nature. An overdue rethinking and reconfiguration of humanity’s relationship to the natural world requires more self-reflected reason and *Eingedenken der Natur im Subjekt* (“remembrance of nature within the subject”), not less; it requires the realization of more humanity in society and a more humane society, not, to the contrary, the sacrifice of humanity as an idea and practice.

Hence, the eclipse of reason does not only find expression in new forms of social and political authoritarianism, the widespread consumption of and damage caused by digital media such as TikTok, or in humanity’s practical oblivion towards the climate catastrophe. Unreason also translates into regressive forms of philosophy and ideology such as posthumanism, which openly render the idea of humanity superfluous and affirm the dissolution of human subjectivity and morality altogether—and reject the idea of progress in any meaningful way. By contrast, a critical understanding of progress ultimately points to humanity’s still unfulfilled realization and remains deeply indebted to the notion of humanity even in the face of its potential dissolution.

6. Conclusion

I have argued that Adorno’s work continues to illuminate problems of progress and its (conceptual, moral, and empirical) antinomies, recognizing both its entanglement in society and its critical qualities, in the contemporary age of unreason. Today, the idea of progress is confronted by global authoritarian regressions, the looming climate catastrophe, the partial dissolution of recognizable boundaries between human agency

and artificial intelligence, as well as omnipresent forms of digital reproduction in which seemingly every human interaction and reaction is identified, measured, counted, objectified, and valorized. Rather than resigning to a negative telos in light of today's hyperreified digital society and the all too realistic threat of human-made disaster, however, a critical and dialectical notion of (both moral and social) progress that takes inspiration from Adorno's writings upholds the possibility of genuine betterment, of averting catastrophe, and of redemption while reflecting the societal trends and conditions leading to potential destruction.

With Adorno I suggest that without the idea of and attachment to humanity yet to be genuinely realized, there can be no progress, and vice versa. A critical understanding of progress thus ultimately points to both democratization and humanity's still unfulfilled realization, its self-conscious reconstitution as a global subject (Rensmann 2016). Progress, which is inevitably also moral progress, remains deeply indebted to the notion of humanity even in the face of its potential dissolution. Furthermore, no philosophical and moral progress seems possible, and no path to avert societal breakdown towards total irrationality, dictatorship, and ultimately human extinction—including the looming climate catastrophe threatening humanity, animals, and the environment as we know it with eradication—seems realizable without the critical resources of human subjectivity. Adorno reminds us that no moral, social, and political progress seems feasible without advancing toward a better, more enlightened, just, and democratic society providing the condition of possibility of humanity's realization.

In addition to soberly reflecting on powerful oppressive and regressive societal trends undermining the very possibility, indeed the very idea of moral progress and freedom, there is the ongoing need to work toward a positive concept view of enlightenment and of progress liberated from the "entanglement in blind domination." (Horkheimer & Adorno 2002, p. xviii). These are the enduring conceptual and normative lessons we can draw from Adorno in preparation of a Critical Theory of progress in and for our time.

Bibliography

- Adorno, T.W.
 1973 *Negative Dialectics*. New York: Continuum.
 1998a Progress. In: Adorno, Theodor W., *Critical Models. Interventions and Catchwords*, New York, Columbia University Press, pp. 143-160.
 1998b Critique. In: Adorno, Theodor W., *Critical Models. Interventions and Catchwords*, New York, Columbia University Press, pp. 281-288.
 2006 *History and Freedom: Lectures 1964-1965*, Cambridge: Polity Press.
 Basnett, Caleb J.

- 2023 Adorno's Cosmopolitan Solidarity, *New German Critique* 50, 1: pp. 31-58. <https://doi.org/10.1215/0094033X-10140719>
- Bennet, L.
- 2016 Thinking like a Brick: Posthumanism and Building Materials, In: Taylor, C.A., Hughes, C. (eds), *Posthuman Research Practices in Education*, Palgrave Macmillan, London. https://doi.org/10.1057/9781137453082_5
- Braidotti, R.
- 2013 *The Posthuman*. Cambridge, Polity Press.
- Bueno, A.
- 2021 What Comes After Depression? The Crisis of Neoliberal Subjectivity and the New Authoritarian Wave in Brazil, *Krisis: Journal for Contemporary Philosophy* 41, 1: pp. 45-64. <https://doi.org/10.21827/krisis.41.1.37167>.
- European Parliament
- 2024 Artificial Intelligence Act: MEPs adopt landmark law, 13 March, <https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law>
- Ferrando, F.
- 2016 Humans Have Always Been Posthuman: A Spiritual Genealogy of the Posthuman. In: Debashish, B. et al. (Eds.), *Critical Posthumanism and Planetary Futures*, New York, Springer, pp. 243-256.
- Gandesha, S.
- 2018 "Identifying with the aggressor": From the Authoritarian to Neoliberal Personality, *Constellations* 25, 1: pp. 147-164. <https://doi.org/10.1111/1467-8675.12338>
- Haraway, D.J.
- 2016 Staying with the Trouble: Anthropocene, Capitalocene, Chthulucene, In: Moore, J.W. (Ed.), *Anthropocene or Capitalocene?* Oakland, PM Press, pp. 34-77.
- Hoffman, S.
- 2022 China's Tech-Enhanced Authoritarianism. *Journal of Democracy* 33, 2: 76-89.
- Hofstätter, A.
- 2019 Catastrophe and History: Adorno, the Anthropocene, and Beethoven's Late Style, *Adorno Studies* 3, 1: pp. 1-19. <https://www.adornostudies.org/wp-content/uploads/2020/11/hofstater-final.pdf>
- Holma, K., Huhtala, H-M.
- 2016 Non-idealizing the Theory of Autonomy: Theodor W. Adorno's Psychological and Political Critique of Immanuel Kant. In: Moses, M.S. (Ed.), *Philosophy of Education 2014*, Urbana, IL, Philosophy of Education Society, pp. 373-381. <https://helda.helsinki.fi/server/api/core/bitstreams/3dd45bcd-ba31-47e4-a20b-4ca6a59c4fea/content>
- Horkheimer, M., Adorno, T.W.
- 2002 *Dialectic of Enlightenment. Philosophical Fragments*, Stanford, CA, Stanford University Press.
- Jäger, D.
- 2021 Affirmativer Charakter heute, In: Sünker, H. (Ed.), *Theodor W. Adorno: Aktualität und Perspektiven seiner Kritischen Theorie*, Münster, Verlag Westfälisches Dampfboot, pp. 176-198.

Rensmann, L.

2016 National Sovereignty and Global Constitutionalism: An Adornian Cosmopolitan Critique, *Critical Horizons* 17, 1: 24-39. <https://doi.org/10.1080/14409917.2016.1117811>

2017a *The Politics of Unreason: The Frankfurt School and the Origins of Modern Antisemitism*, Albany, NY, SUNY Press.

2017b *Critical Theory of Human Rights*, In: Thompson, M. (Ed.), *The Handbook of Critical Theory*, New York, Palgrave Macmillan, pp.631-653.

Sorgner, S.L.

2021 *We Have Always Been Cyborgs: Digital Data, Gene Technologies, and an Ethics of Transhumanism*, Bristol, Bristol University Press.

Timeto, F.

2021 Becoming-with in a Compost Society — Haraway beyond Posthumanism, *International Journal of Sociology and Social Policy* 41, 3/4: pp. 315-330. <https://doi.org/10.1108/IJSSP-08-2019-0158>

Yang, X.

2019 China: Wir sehen Dich! *Die Zeit*, 10 January, https://www.zeit.de/2019/03/china-regime-ueberwachungsstaat-buerger-kontrolle-polizei?utm_referrer=https%3A%2F%2Fwww.google.com%2F.

Zuboff, Sh.

2019 *The Age of Surveillance Capitalism*, New York, Public Affairs.

*Michele Vadilonga Gattermayer**

From the Anthropological Machine to the Animal Crisis

Abstract

In this essay, I will reconstruct the biopolitical framework regarding human-animal relations, through the device of boundary production Agamben called the “anthropological machine”. I will argue that the effects of the machine can be more extensively grasped on the constitutive outside of the human, namely animals and that the current status of institutionalised relationships with animals, gathered under the name of “animal crisis”, can be framed as a direct effect of the separation.

Keywords

Anthropological Machine; Biopolitics; Non-human Animals; Human/Non-human Divide; Ecofeminism

1. Introduction

In this article I will reconstruct the bio-political framework regarding human-animal relations, firstly inquiring into the device Agamben called the “anthropological machine”: the ontological device of boundaries production and reproduction. To clarify the practical and historical functioning of the machine, I will identify two different sub-devices that serve as a sacrificial strategy of protection of “the human” against “the animal”. I will then argue that the effects of the machine, contra Agamben who exclusively focuses on the category of “the human”, can be more extensively grasped in their complexity on the constitutive outside of the human, namely animals. My claim will be that the current status of institutionalised relationships with animals, which involves practices criticised as violent, oppressive and unjust from animal ethics scholars,

* SISSA Trieste.



can be framed as a direct effect of the anthropological machine, gathered under the name of “Animal Crisis”. The aim of the article is not to draw an automatic moral or political condemnation of these practices, but to clarify the deep link between the bio-political separation between humans and animals and the animal crisis. The contribution that the bio-political framework offers to this discussion is, I argue, central, since it put at the forefront that the human/non-human boundary is not only a “product of detached reason”, nor simply a reaction to special feelings as those Steiner (2024) calls “pathos of fear and arrogance”, but rather the consequence of a political process of exclusion the branches of which form our present.

2. The Anthropological Machine

In the history of Western philosophical and scientific thought, animals have been intensely present as objects of study of particular relevance, since their characteristics shared with humans (firstly mobility, a spectrum of sensitivity, and abilities such as vocal communication, social structures, empathetic behaviours and many more) pose a challenge to the status of human beings. According to Piazzesi (2017, p. 161), “in front of them we experience an asymmetrical mirroring that evokes in us a mixed feeling of familiarity and monstrosity, which dooms them to an uncertain status with regard to *our* identity”¹. The “definitional anxiety”, generated by the necessity of creating a precise definition of “what is a human” with clear and sharp boundaries, made the animals at the same time privileged and troublesome objects of study. The development of new scientific fields only led to a perpetuation of this “uncertain status”, showing that the challenge they pose to the preservation of a closed “human identity” was quite difficult to erase.

Aprioristically excluded from the subjective knowledge of what were to become the “sciences of spirit” [*Geistwissenschaften*] first and the “cultural sciences” later, but difficult to incorporate at the same time, as living things, into the objective knowledge of the natural sciences, animals represent the uncomfortable guests of modern episteme (Piazzesi 2017, p. 162).

The reason behind this wobbly condition is that animals play a central role in the system of exclusion which still persists in our contemporary world, embedded in the political and societal structures of Western

¹ All translations from Piazzesi 2017, Mormino 2017, Colombo 2017, Pellegrino & di Paola 2018, and Pollo 2021 are made by the author.

industrial and post-industrial societies (and also in many non-Western cultures such as China and India). The human/non-human divide has been historically produced and reproduced through the severing between humans and animals, and many of the strategies and mechanisms of maintenance of the divide have used the human/animal boundary as a focal point. One of the foundational and most important ways of structuring and perpetuating the divide through centuries is the one Agamben (2004) called the “anthropological machine”, a powerful device continuously producing and establishing the division between the human and the animal.

Agamben reconstructs the functioning of the fundamental and enduring division between humans and the other parts of reality, identifying the place of the most original severing in the human/animal divide. The broader boundary between humans and non-humans is produced through this narrower divide: if we manage to justify the sharp boundary between human beings and animals (especially the most similar to us) and to build on this boundary hierarchical forms of political dominion and social structures, *a fortiori* the scope of the exclusion can be extended to the enormous realm of plants, fungi, bacteria, all other organisms belonging to the tree of life and non-organic entities: the non-human.

The turning point for the origins of the human/animal divide in the Western tradition is placed by Agamben in the Aristotelian isolation of nutritive (or vegetative) life, as the most general identifiable way of “being alive”. This strategical isolation is useful for Aristotle to structure the division and hierarchy between all the different forms of life, which are thus separated in virtue of their gradually higher faculties. This foundational division which made possible all other separations (vegetative/relational life, organic/animal, animal/human) is internal, it passes within the human and at the same time produces it: “What is man, if he is always the place—and, at the same time, the result—of ceaseless divisions and caesurae?” (Agamben 2004, p. 16). The effort of the Italian philosopher is thus to investigate the practical and political consequences of this separation, highlighting how, through history, the device of boundary production articulated different strategies, sometimes with terrible outcomes for both sides.

Agamben describes two variants of the machine: the one of the moderns, and the one of people of “earlier times”. Their functioning and results are homologous, but they structure the terms of the divide in two different ways. Both the machines work through a radical exclusion of some elements, and an inclusion of others, to protect and affirm the category of “human” in opposition to what is not human (the animal, or the inhuman). Since the purpose of the machine is to safeguard one of the terms of the opposition, its functioning is rigged, as “the human”

is an already presupposed category: “the machine actually produces a kind of state of exception, a zone of indeterminacy in which the outside is nothing but the exclusion of an inside and the inside is in turn only the inclusion of an outside” (Agamben 2004, p. 37). The ancient variant of the machine worked through the inclusion in the category of “human” of some elements of the constitutive outside via a “humanisation of the animal”: those who were found in the zone of indeterminacy were captured inside the human sphere as “messengers of man’s inhumanity, the witnesses to his fragile identity” (Agamben 2004, p. 30). The *enfant sauvages* (individuals who, from a very young age, were raised apart from other humans and never developed human social skills such as language, social behaviour or care) and the ancient figures of the slave and the barbarian, seen as animals in human form, are included among the protagonists of this variant of the machine. The anthropological machine of the moderns works differently: it produces the outside through the exclusion of elements of the inside. It operates an “animalisation of the already human”, isolating the animal already inside the human life (this is reminiscent of the isolation of the vegetative life from which the functioning of the machine has begun). The feverish research of the late XIX century for the “missing link” between humans and apes, after the acceptance of the Darwinian theory of evolution of all species (humans included), is for Agamben a paradigm of the effort of finding a marker of the divide, making it tangible and incontestable. An element he emphasises is that, in contemporary times, the modern variant of the machine showed its brutality and dangerousness when, in lieu of “this innocuous paleontological find we will have the Jew, that is, the non-man produced within the man” (Agamben 2004, p. 37)². While the machine of the ancients captures those found in the zone of indeterminacy inside the sphere of “the human”, its modern variant excludes them in virtue of their “inner animality”.³

² An extremely contemporary example of the same “animalisation” can be found in the words of the Israeli Defense Minister who defined “human animals” the Palestinians inhabitants of the Gaza Strip, while announcing the complete siege of Gaza on October 9, 2023: “We are fighting against human animals, and we are acting accordingly”. (Fabian 2023)

³ A parallel theory about the human/natural divide (or dualism) and its deep connection with modernity is Bruno Latour’s idea of the “Constitution of Modernity” (Latour 1993), which prescribed the strict division between the human and the non-human world, the social and the natural world. The Constitution ultimately crumbled on itself since, while the modern narrative prescribed a strict dualism, a parallel modern practice continuously produced hybrids of the two worlds, which were, with time, impossible to contain while holding the Constitution valid. While Latour’s idea is more focused on the process of modernisation and its sociological, anthropological and political meaning, it harmonises with some insights of Agamben’s bio-political theory (I thank the anonymous reviewer for this point).

Agamben does not go further in detailing the practical functioning of the machine, he only schematises the fundamental ontological structure it operates to produce and protect the closed category of “the human” through the ontological and constitutive exclusion of the animal from it. I argue that we could identify two “sub-devices” historically employed by the machine in Colombo’s discussion on human dominion over animals: the theory of “*Homo duplex*”, and human exceptionalism. *Homo duplex* is the name of the view, largely shared in history by philosophers and scientists, that describes humans as organisms with a double nature: a material/animal one, and a spiritual/human one. Colombo (2017, pp. 99-105) reconstructs this view through the writings of the XVIII century naturalist Buffon, who openly speaks of this double nature in his *Histoire naturelle*. The belief in the existence of these two “spirits” inside each human individual, which are often in clash with one another, generates the narrative that to be “fully human”, in the sense of fulfilling in the best way possible the human nature, it is a duty to defeat and abandon the internal “non-human” element, an idea that, indeed, dates back to the greek division of the soul. The greatness of humans, thus, would be in leaving behind, eliminating a part of themselves, to achieve full mastery of properly human faculties such as reason, freedom and language. The fracture between the human and the animal is internal, as Agamben argued: it is a constitutive element of the “human nature”. The device of *Homo duplex* is coherent with what we have called the “animalisation of the human”: to operate the divide, the machine excludes a part of the inside (the human) in virtue of its inner animality (or inhumanity). This research of the “animal” inside the “human” is a strategy historically used to establish hierarchies and structural exclusions and oppressions to the detriment of large numbers of humans who do not enter the “properly human” paradigm. To set the threshold too high is to exclude many humans from the concept of “humanity”, in virtue of the lack of the relevant character. This has historically marked the creation of normative ideas of humanity (prescriptions of what a human is to be) from which a large number of humans could be, conveniently with the power structures in place, politically oppressed and excluded: racism, the exclusion of women from political and social life, of non-neurotypical and disabled people, and peculiar cultural or religious groups have found a powerful ally in these normative and exclusionary ideas of humanity. As Colombo (2017, pp. 101-102) argues:

the idea of *homo duplex* contributes to clarify man’s self-identity only at the cost of a continuous hierarchisation that affects both the individual and humanity itself. [...] [T]o set a purely natural or material dimension that must be overcome for the achievement of a higher dimension inevitably creates

hierarchies between children and adults, as well as between different individuals and peoples.

In the *Homo duplex* narrative, widespread and largely adopted and accepted still today, it is already embedded the idea of “human exceptionalism”⁴: from the claim of the existence of these two heterogeneous natures, it follows that there is something, some element, some capability, that is “uniquely human”, and that serves as a boundary. The divide has been historically marked in different ways, and many different strategies of justification have been deployed, isolating one or more characters as the signs of “human exceptionality”. One of the most enduring has been the idea of a *scala naturae*, a natural hierarchy that placed humans ontologically above animals (and thus also above plants and inorganic entities), because of their proximity to God, who created humans (and only humans among material beings) in his own image. Humans are exceptional, among material beings, because equipped with a spiritual principle which, providing them the capacity for language, reason and love, brings them closer to God, and justifies the animal servitude to them. This fundamental ontological ranking of beings, developed by Greek Neoplatonism and Christian Neoplatonic scholars, provided an image of humans as “suspended” between the spiritual world of perfection, and an animal material world of death and impurity, as in the theories of Pico della Mirandola. In this way, the divide between humans and animals was defended through a divine spiritual principle which was uniquely proper to humans. With the progressive secularisation of culture and science the “spiritual principle” was substituted by other

⁴ The concept of human exceptionalism has a long and branched out history, which led to an idea of humanity as radically separated from its natural environment. One of the branches of its history (and one of the most long lasting and persuasive) can be identified in the Promethean myth. This narrative describes human history as a progressive emancipation from Nature: from a world dominated by necessity and instincts, death and suffering, toward the erection of a parallel world completely designed by human hands, where culture, religion, morality, technology and arts could emerge – as the gift of fire by the titan symbolises (Aeschylus, vv. 108-111). The history (and perhaps the destiny) of humankind is, for the Promethean narrative, one of ascent through emancipation from Nature, toward an Earth shaped in our image. According to some scholars, the concept of Anthropocene (as the proposed name for the current geological epoch, marked by the extensive and global human interference with the planet’s biological, chemical and geological cycles) marks the moment in which humans and nature definitively part company – according to eco-modernists, for the better, since the ultimate source of the ecological crisis is the “over-reliance” on ecosystems (Asafu-Adjaye et al. 2015). The Anthropocene is, for this narrative, the apex of our species’ exceptionalism, the evidence of our superior capability of managing (for the better or for the worse) the entire planet, changing its history. Either a God, or a cancer, humankind is no longer a natural entity because of its exceptionalism.

characteristics such as language (as in Hobbes), reason (as in Descartes), cooperation (as in Durkheim), and many more, creating the effort for the defence of a boundary which was progressively more unstable the more the knowledge of animals and nature evolved. Darwin's revolution in the mid XIX century has been a turning point for the discourse on "human exceptionalism" because of the scientific assertion and assessment of the common ancestry of all animals and their complex patterns of evolution and modification from previous species. After the acceptance of Darwin's theory of evolution, to claim the radical heterogeneity of the human nature from the animal was impossible, since it meant that the boundary between humans and their closest "cousins" was historical, not ontological, going back to the last common ancestor between humans and their closest living relatives: chimpanzees and bonobos⁵. With the discovery and study of other extinct species of *Homo* that predates the advent of *sapiens*, the idea of placing a close boundary between human and non-human life gets even more difficult.

Homo duplex and human exceptionalism are sophisticated and varied instruments deployed by the anthropological machine to continuously produce and reproduce the boundary, which "decides each and every time in favor of the human" (Broglia 2013, p.1). The interior nature of the divide shows its being a foundational human fact, constitutive of the category of "the human", and consequently, in opposition, of "the animal".

3. *Homo nosce te ipsum*

Agamben highlights how the anthropological machine emerged in its fundamental structure in history recalling, among other examples, Linné's work, the founder of scientific taxonomy. More than a century before Darwin, Linné already expressed strong doubts about the sharp boundary between humans and apes defended by the mainstream culture of his time. His observational work on apes led him to neatly reject Descartes' philosophy that relegated the animals to the position of *automata mechanica*, efficient self-propelled machines lacking of *res cogitans* (the metaphysical substance of reason, feelings and language), with the famous ironic sentence "*Cartesius certe non vidit simios*" (Agamben 2004, p. 23). In light of this challenge, and the refusal to accept any non-scientific justification for the divide, in his *Systema Naturae* he inscribes the genus *Homo* in the *Anthropomorpha* order, together with the genus *Simia*,

⁵ It is estimated that the last common ancestor between the *Homo* and the *Pan* (chimpanzees and bonobos) genuses lived around 6 million years ago (Patterson *et al.* 2006).

Lemur and *Vespertilio*. The neat inclusion of humans among the group of Primates (and, quite interestingly, bats) is not the most interesting act of the Swedish naturalist, according to Agamben. He focuses on the name chosen for our own species: *Homo sapiens*. Differently from all other species, which are described with a bimodal nomenclature (genus + species: *Pan troglodytes*, *Canis aureus*), the name of our species “*sapiens*” is not a description, but rather a trivialisation of the philosophical maxim “*nosce te ipsum*”: “know thyself”⁶. This meant, for Linné, two things: first, that it was not necessary to explain with a specification what a human was, since each one of us is able to “know themselves”, to recognise themselves as humans; second, that the “specific identity” of humans could be exactly this, to be able to recognise themselves. Agamben summarises this conclusion: “*man is the animal that must recognize itself as human to be human*” (Agamben 2004, p. 26). *Homo nosce te ipsum*, shows, for Agamben, that embedded in the effort to classify humans there is the anthropological machine. When we try to focus our sight on humans and divide them from animals, we activate this “optical machine constructed of a series of mirrors in which man, looking at himself, sees his own image always already deformed in the features of an ape” (Agamben 2004, pp. 26-27). The product of the machine is the human itself, which stems from the exclusion of the outside.

This ontological operation is for Agamben foundational of Western politics and philosophy and shows that the divide is a continuously reproduced political decision “in which the caesurae and their rearticulation are always dislocated and displaced anew” (Agamben 2004, p. 38). The divide is thus essentially and internally human, an ontological and political operation maintained through time with many different but structurally similar strategies. We can conclude then that the unveiling of the anthropological machine exposes the true nature of “human exceptionalism”, since the “uniquely human”, “this ‘something more’, however, seems to be the desire to draw a boundary and establish an exception” (Colombo 2017, p. 110).

Many authors, especially in the field of human-animal studies (Broglio 2013; Mengozzi 2021), have highlighted the fact that, in his discussion of the anthropological machine, Agamben is only concerned with one side of the machine, “the human”. The main interest of the Italian author, in line with his overall work, is to examine the consequences of the exclusion of parts of humanity from humanity and “humanness” itself (a case in point is that of the Jews). This privileged focus on humans is evident since almost half of the book is devoted to an in-depth dis-

⁶ The name *Homo sapiens* has been used from the tenth edition of the *Systema Naturae* of 1758.

discussion about the Heideggerian concepts of “the Open” and the animal “poverty-in-world”. Animals in their variety and internal complexity lie in the background of this discussion, they are relegated to the position of a unifying category, an “empty signifier” (Laclau 2003). Animals are, in Agamben’s discussion, an undifferentiated constitutive outside, their role being merely a means of production of “the human”. The internal complexity and differentiation captured under the signifier “the animal” are left aside by the philosopher, perpetuating a common trend of construction and essentialisation of a completely artificial category. When we talk about animals, especially in discourses concerning animal rights, human-animal relations, animal welfare and other important issues, we often limit our concern to the animals closer to us, or with whom we have more familiarity (primates, mammals, pets, sometimes fishes and reptiles, and more in general vertebrates), but we rarely have in mind the whole internal complexity of “animals”⁷.

I will now examine the consequences of the anthropological machine on its outside, the “excluded par excellence”, namely non-human animals, in their practical and political dimension, arguing that the origins of the practices to which we expose animals on a daily basis there is the ontological (and physical) production and unilateral control of the animal body with the objective of producing, in a privative way, the human. This foundational ontological operation shows that, as Agamben (2004, p. 80) claims, “in its origin Western politics is also biopolitics”, and that the fundamental conflict that permeates our cultures is between humans and animals, categories produced and reproduced by a sacrificial mechanism of radical exclusion.

4. The war against animals

One of the voices who better depicted this fundamental conflict with clarity and disenchantment is Thomas Hobbes (2012), who famously described the relationship between humans and animals as a “perpetuall warre”. In war positive right has no place, the only right is the natural right: this is the condition of the state of nature, where there is no commonly recognised sovereign authority enforcing positive law and the only right of each individual is an *‘ius omnium in omnia’*. While, famously, human beings, thanks to their calculative reason and natural laws, are able to create a “social pact” and enter a civil society, with non-human animals, Hobbes wrote, this is not possible, since their lack of reason

⁷ It is unlikely that, while talking about “animal issues”, we refer to arthropods (insects and arachnids), echinoderms (starfishes and sea urchins), cnidarians (jellyfishes and corals) or other members of the 40 known animal *phyla*.

and language make it unfeasible to make a pact and enforce compliance⁸. Without a common authority to whom confer all natural rights, human-animal relations cannot exit a state of nature, a condition of perpetual war of all against all, in which the only right is strength, and any act of violence can be performed with impunity. Hobbes' merit is the severing from the enduring tradition of Christian theology that founded human dominion over animals on positive divine right. Against this view, he unveils with great lucidity that the dominion stems from the mere natural right of strength: we do what we want to animals because we can, and because we have no reasons to fear them.

The condition of war pictured by the modern philosopher is harsh and violent, but interestingly symmetrical: "Forasmuch therefore as in proceeds from the right of nature, that a beast may kill a man; it is also by the same Right, that a man may slay a beast" (Hobbes 1983, p. 87). However, the war against animals has historically been "distinctly asymmetrical and not only because we were the ones who started it: the disproportion of the forces on the battlefield and their different organization are in fact immeasurable" (Mormino, Colombo, Piazzesi 2017, p. 9). Humans' relationship with animals was one of symmetrical predation for thousands of years, but structurally transformed starting from the Agricultural Revolution. Following Mormino, I call "dominion" the asymmetrical condition of control and violence toward many non-human animals starting with sacrifices and domestication. The relevant shift is from a practice of catching/killing wild animals, to "bringing the animal into existence, in ways and at times useful to humans, and the exploitation of its life processes in their entire duration" (Mormino, Colombo, Piazzesi 2017, p. 7). Human military superiority and organisation led to a condition of systematic control over many species of animals, relegated to a situation of perpetual slavery: they do not have a say on their exploitation and the unidirectional dominion over them, they are often legally considered "propriety" of their human owners. The war against animals probably began as symmetrical predation but changed radically in an asymmetrical system of sacrificial dominion in the last 12,000 years.

Mormino, Colombo and Piazzesi mention three macro-effects of the war against animals in chronological order: domestication, genetic manipulation and engineering, and extinction. The practice of domestication has started around 12,000 years ago with the enclosure and exploitations of large numbers of specific animals for human-centred purposes

⁸ The kind of "social pact" observed by the social species that Hobbes described as "political" (bees, ants) is completely different in nature. While the human pact stems from calculative prudential reasoning, it is artificial, that of animals is "natural", they "naturally" aim at the same objective and cooperate without contentions (Hobbes, 1983, p. 87).

(probably for symbolic/religious reasons firstly, and later for economic motives) (Mormino 2017): to its “evolution” and “institutionalisation” and the practice of industrial farming will be devoted the last part of this article. The processes of genetic manipulation show the strong biopolitical nature of the structure: the production and modification of animal bodies is material, not only ontological. Billions of non-human individuals are meticulously brought into existence with the only purpose of being “utilised” and killed. The artificial selection of individuals with “better” characters (for human needs and desires) has marked the transformation of numerous species: farm animals have been selected in order to produce more meat, milk or wool, and to be more tame and docile; animals used in laboratories have been engineered to develop certain specific characters useful for the research (the “oncomouse”, a mouse whose genetic code has been modified to increase its susceptibility to cancer, is a famous example). Also, animals considered “pets” have a long history of genetic manipulation: domesticated dogs not only lost many characteristics of their wolf ancestors but experienced feverish selections that shaped the great variety of dog breeds, so different from each other. The example of dogs sheds light on one of the most cruel and violent consequences of the process of modification: the creation of “harmful phenotypes”, namely individuals whose selected characters generate great suffering or early and inborn pathologies⁹. We have canalised the evolutionary process of many animals in directions that, without our interference, would have been impossible, sometimes deliberately, other times with little understanding, sometimes through invasive methods of DNA engineering, other times with simpler artificial selection of mating individuals. Humans have practised genetic modification for different purposes: economic (breeds of cow producing up to 10,000 kg of milk per year)¹⁰, scientific (the oncomouse), aesthetic (the pug) and even artistic (as the bio fluorescent rabbit Alba “created” by Eduardo Kac) (Harari 2014, pp. 402-403). At the expense of the already excluded, genetic manipulation modified the physical conditions of many species, harmlessly in the best scenarios, with a great amount of violence and pain in others.

⁹ One of the most evident examples is brachycephaly: common among many different dog breeds (bulldogs, pugs, chihuahuas, Cavalier Kings, Dog de Bordeaux), this condition stems from the selected characters making the head of the dogs smaller and flattened for aesthetic reasons and generates severe health problems to the respiratory and digestive systems, to the spine, ears, eyes and skin. This issue is mentioned by Pollo (2021, pp. 90-91).

¹⁰ This is the case of the Holstein Friesian cow. Similar cases are the selection of sheep that produce such an abundant amount of wool that they cannot survive without being sheared or that of chicken breeds such as the broiler who reach 1,5 kg of weight in just 30 days of life (they are another harmful phenotype, prone to cardiovascular and skeletal dysfunctions).

Species extinction is the third consequence of the war against animals, which emerged as a concerning issue in the last decades. The complete destruction of other species is not a new fact in human history: it has been suggested that behind the extinction of other *Homo* species such as *neanderthal* or *denisova* there could be the hand of the first groups of *sapiens* that reached Europe around 40,000 years ago, also through deliberate actions of warfare, aggression and systematic killing, in a quasi-genocidal fashion (Harari 2014, pp. 23-24). Also, it has been hypothesised that the extinction of the so-called “Pleistocene megafauna”, the group of many big size species (among which mammoths and giant sloths) disappeared in the last half of the Late Pleistocene (50-12,000 y. a.), could have been con-caused by the migration of *Homo sapiens* groups in new regions¹¹. Destruction of habitats, pollution, extensive urbanisation and climate change are causing the disappearance of animal species at an impressive rate, leading to what has been called “Sixth Mass Extinction”, an ecocide of immense dimensions: today, the rate of extinctions has increased exponentially, up to a thousand times higher than the “background rate of extinction” (Pimm *et al.* 2014).

The convergence of the practices of domestication, genetic manipulation and species extinction generated what Crary and Gruen (2022, p. 14) called “Animal Crisis”: a situation in which “human-animal relations have reached a desperate point”. Biodiversity loss and the systemic killing of animals on an industrial scale reinforce the relation of oppression and dominion over non-humans, whose lives are day-by-day exploited and commodified. According to these authors, “Anthropogenic animal destruction also includes the deliberate creation, exploitation, and killing of animals in laboratories, hunting grounds on land and in the oceans, aqua-farms, and land-based industrial farms” (Crary & Gruen 2022, p. 12). In terms of mammal biomass, wildlife makes up only about 4%, while livestock amount to 62% (humans form the remaining 34%), and, as scientific studies show, in term of biomass “humans and livestock overweight all vertebrates combined, with the exception of fish” (Bar-On, Phillips, Milo 2018, p. 6508)¹². To this proportion, we should integrate another data: that of animal killing. According to FAO data (Roser 2023), more than 200 million cows, pigs, chickens, goats, sheep and ducks are slaughtered for meat and dairy every day (this data excludes millions of fish, the number of which is difficult to estimate). That means that, as the

¹¹ The role of humans in the Quaternary extinction event is debated among scientists. However, the coincidence of human migration in Australia around 50,000 y. a. and in the Americas around 13,000 y. a. with the extinction of local megafaunas is a strong clue (Sandom *et al.* 2014).

¹² The situation is homologous for what concerns domesticated and wild birds.

Humane Foundation calculated, “if humans killed each other at the same rate we kill animals, [...] we’d be extinct in 17 days” (Crary & Gruen 2022, p. 13)¹³.

5. The system of dominion

The ecofeminist tradition has always regarded the contemporary system of animal oppression and dominion as focal point of discussion, highlighting how structural and systemic exploitation and destruction of animals and nature are rooted in the same framework that oppresses and subjugates “women, the poor, colonized, racialized, and other marginalized people” (Crary & Gruen 2022, p. 130). The ecofeminist framework is grounded on the critique of all hierarchical oppressive relations and divides (man/women, white/non-white, human/animal, nature/culture) which are interpreted as “structurally interrelated” (Crary & Gruen 2022, p. 3). The human/non-human divide can be interpreted as the first form of exclusion, from which all other exclusions were shaped, among which hierarchical structures of oppression such as patriarchy, monarchies, animal exploitation and non-criminal killing: their origin is the sacrificial exclusion of a constitutive outside. Ecofeminist scholars put emphasis on the elements of connection between practices of women and animal oppression, such as the “reproductive and sexual enslavement of female animal bodies” (Gaard 2012, p. 524), the role of violence in the maintenance of such systems, the feminisation of feminist empathy and care for animals’ oppression, the link between the practice of eating meat and male supremacy (Adams 2015). From this standpoint, many practices of human dominion over animals can be described as fully-fledged forms of slavery. The fight against commodification, violent coercion, exploitation and killing of non-human animals is regarded by ecofeminists as a structural part of the common and intersectional struggle for liberation and recognition. The deconstruction of the system of oppression and exclusion is for ecofeminists the primary goal of animal ethics and politics.

Many practices of human-animal relationship are mentioned as forms of “animal slavery”. Simone Pollo (2021, pp. 85-92), for example, de-

¹³ This calculation must be interpreted in light of the immense number of non-human animal individuals (especially livestock) compared to the relatively small number of humans (8 billion individuals). According to FAO data, livestock chickens amounted to more than 21 billion in 2014, while cows and other cattle species to 1,5 billion: Our World in Data, Chart: Livestock counts, World, 2014, <https://ourworldindata.org/grapher/livestock-counts?time=latest>.

scribes in his Manifesto activities which he deems incompatible with the ethos and the principles of liberal democratic societies: the first he mentions is the usage of animals in circus performances and other similar shows, such as in dolphinariums where dolphins, whales and orcas perform in “ability tricks” and “comic shows” with their human owners. The animals involved in these practices are, almost always, individuals of wild species (whose needs are often incompatible with an enforced co-habitation with humans), they often undergo a stressful and violent process of taming, learning actions and commands completely out of tune with their wild behaviours and routines. Other forms of animal slavery described in the Manifesto are *corridas*, sport hunting, the intentional selection and trade of harmful phenotypes and fur farming. While these practices are deemed incompatible with democratic societies, and thus democratic animalism is committed to a complete abolition, Pollo argues that for other activities, such as animal experimentation for scientific and health research, democratic animalism “must necessarily stand on a ‘reformist’ position, that is, one oriented toward advocating for ever-increasing protections for the welfare of animals involved in experimentation and for making it a political and scientific priority to find alternatives *to* and *within* animal experimentation” (Pollo 2021, p. 82).

Interestingly, one of the practices that is mostly discussed in animal ethics debates, and that many scholars deem as the most paradigmatic instance of animal slavery, is absent in Pollo’s list: animal farming in its contemporary industrial apparatus. This eminent exclusion is defended through a discussion on the cultural, symbolic and traditional meaning of the practice of animal farming and eating meat. Domestication and farming are the first instances of dominion over animals, they have accompanied the development of human civilisations since the construction of the first cities, being a fundamental element of the Agricultural Revolution, an event (or better, a series of parallel events) which marked a radical and irremediable transformation of the ways of life of many human groups: “human civilisation is structurally founded on the use of animals” (Pollo 2021, p. 63). In another essay, Pollo (2017, p. 148) stresses this point against different theories defending vegetarianism as a moral universal imperative stemming from the recognition of animal suffering and exploitation. He argues that these accounts, such as the classical animal rights theories of Singer and Regan, “underestimate the importance and complexity of meat-eating habits in the human form of life”. Animal domestication and farming for meat, milk and wool are foundational elements of (at least) Western societies: these practices stem from and crystallise the human/non-human divide in the form of the human/animal separation. But to state that these institutionalised

actions are part of the deep structure of our societies is not at odds with claiming their being oppressive, violent and exploitative practices.

The argument Pollo uses to support this idea works on two binaries: the first is the deep cultural and traditional element just discussed, and the second is the issue of “unnecessary harm”. Circuses, dolphinariums, fur farms, and other activities deemed forms of animal slavery, are considered by Pollo unjustifiable because they cause harm for frivolous reasons: they sacrifice the well-being and dignity of the animals involved for unnecessary purposes, perpetuating in this way “the anthropocentric idea that animals are a commodity always available for human interests and desires” (Pollo 2021, p. 90). I would argue this reasoning underestimates two elements: first, the fact that the possible cultural and traditional importance for human societies of those practices (some of which have, such as circuses and sport hunting, an important and enduring part in human cultural history) are considered “irrelevant”, “unimportant” and “inconclusive” while those of animal farming and meat-eating are regarded as prominent. Even if the latter has a historical foundational role, its cultural value is exaggerated in comparison with the others, the history and cultural values of which are ignored. Second, if their “unnecessary” nature is enough to consider these practices unjust forms of slavery, the same could be argued for animal farming and meat-eating. In contemporary industrial societies, eating meat and using other products originating from animals cannot be described as a “necessity”: we could clearly survive with plant-based or synthetic alternatives. On the contrary, many authors (Burgat 2022, p. 455) showed that “it was precisely at a time when humanity could have stopped using ‘animal resources’ that it instead institutionalised, and therefore generalised and intensified, a murderous relationship to animals”.

Data show that the world consumption of meat grew from 62 million tons in 1950, to 437 million tons in 2011: an increase of over 700% (Larsen, Roney 2013). The animal industry exploded and institutionalised as an integral part of everyday life in contemporary societies in the second half of the XX century, generating a new dimension of life conditions for the many species involved, trending toward a deepening crisis of relationships. We can speak of a fully-fledged “Animal Crisis” only considering these “disruptive innovations”. According to Krzykowski (2021, p. 80), the industrialised process of killing animals led to what he calls “de-animalisation”, which appears to be “the tragic final stage of the consecutive processes of the domestication and instrumentalization of animals in the modern era”, which made them invisible to society, and their bodies inanimate pieces of matter, disposable to be created and processed in an ever-ending industrial cycle. While, before the advent of hyper-industrial apparatus, animals were considered machines, efficient

means of production, objects and instruments humans were free to use and treat as they needed and pleased, the situation changed with the development of global capitalism: “if industrial animals, such as horses or cows [...] could have been seen as *lumpenproletariat* at the beginning of the twentieth century, they have *massively* become ‘quickly used and replaced objects of consumption’ since the 1950s” (Krzykawski 2021, p. 79). Inside the industry, animals have been “de-objectified”, “their bodies became disposable and processable” (Krzykawski 2021, p. 77), ready to be transformed and presented as something “other than itself” in the form of varieties of meat.

An aspect highlighted and discussed by different authors is the strong link between the practice of eating meat and the mechanism of the divide. Derrida (1991) famously used the term “sacrificial structure” to describe the existence of a deep system of production of the subject that includes as a core element the “noncriminal putting to death” of animals. This structure is at the origins of multiple systems of domination and oppression, of the schema he called “*carno-phallogocentrism*” (Derrida 1991, p. 113)¹⁴, and is central for the production of a peculiar contingent subjectivity. The carnivorous virile subject emerges as the production of the sacrificial structure as the ultimate result of the process of systematic exclusion and elimination of parts of reality from the protected sphere of the properly “human”. Florence Burgat (2022) has taken this concept, trying to better describe the link between eating meat and sacrifice. Starting from the claim that the consumption of meat is not a necessity, but rather a deep desire of humanity which must be taken seriously, she interprets sacrifice in the light of the structuralist tradition. A structure is, she argues recalling Lévi-Strauss’ anthropology, a “primordial fact”, a system we find ourselves in since our birth, that influences and shapes all societal systems of relations. Eating meat is, in this framework, intended as a “superstructure” born to institutionalise and defend a relationship of radical severing and hierarchical violent dominion over non-human animals. The desire to eat meat is nothing but an expression of the desire for “a radical and indefinite separation form ‘animality’ that only the eating of animals can fully achieve. To chew something, digest it and also expel it in the form of excrement is to obliterate it in a way like no other” (Burgat 2022, p. 457). This is why Burgat thinks that a universal complete shift toward vegetarianism would be impossible: the desire for meat will persist in any case, since it is deeply rooted in a sacrificial structure of division. What meat-eaters “love” about meat is not, she argues, its mere

¹⁴ This close link between the practice of eating meat (interpreted as a fundamental productive process of virility) and the patriarchal structure has been interestingly developed simultaneously and independently by Jacques Derrida and Carol Adams (2015).

physical substance, but the fact that it comes from an animal. While it could seem at first that this thesis is at odds with the concept of “de-animalisation” presented before, I argue that they are both expressions of the primordial sacrificial mechanism which enable us to watch the same entity (animals) in two apparently contradictory ways. We hide the fact that what we eat is processed animal flesh, disassociating it from its original referent, and at the same time deeply incorporate the subjective necessity of prevailing and eliminating that animal, perpetuating, reproducing, and empathising the radical divide.

6. Conclusion

The advantage of the biopolitical/ecofeminist frameworks deployed in this article is that it is able, as we discussed, to detect the central role played by the human/non-human divide in the formation of the “Animal Crisis”. The divide emerges in being is not only a “product of detached reason”, nor simply the reaction to an affective dimension of fear and arrogance, but rather the consequence of a political process of exclusion the branches of which form our present.

Classical animal rights theories, such as Singer’s or Regan’s, apply “extension strategies” (Pellegrino & di Paola 2018, pp. 137-154): starting from the idea that there exists some characteristics that are bearers of moral value, they support the extension of moral status and rights to all those who are equipped with these traits. For Singer the characteristic giving value to forms of life is the capacity to feel pain and pleasure, whose interests must be protected accordingly, while Regan focuses on subjectivity as the relevant trait. Other theorists focused on agency, sentience, intelligence, conscience and self-consciousness. All these strategies, while praiseworthy for their concern for animal suffering and the strong advocacy for the protection of animal lives, still necessarily operate constituting a protected inside at the expenses of an excluded outside. The expansion of the circle of moral concern can function only if there is someone outside of it. The human/non-human divide is perpetuated in these theories through “the game of ‘how close are you to the [human] model’” (Colombo 2017, p. 131): in virtue of the presence of some arbitrary characteristics, which are always common traits to humans that we value important, a species is included or excluded from the moral circle. The experiments aimed at showing the presence of these characteristics in apes, dolphins, orcas and other species, gained strong popularity, but they are founded on the idea of showing that those species are close to humans, and thus worthy of moral consideration and rights. Those who pass the test are included in the inside because they share something with

us, they are, in part, humans; those who fail it, remain outside. While it is possible to defend the strategic necessity of dividing between different animals while dealing with animal ethics, classical theories are still operating with variants of the anthropological machine, without questioning its structure.

This article is not meant to draw an automatic moral condemnation to the practices described, for which the debate in animal ethics offers multiple arguments, but rather to clarify how the original *cesura* between humans and animals is directly linked to the current state of affairs with its system of almost absolute dominion. This clarification is, in the intention of the author, essential to construct political solution to the Animal Crisis, since it put to the forefront its biopolitical foundation.

Bibliography

- Aeschylus, *Prometheus Bound*, english translation by Ian Johnston.
- Adams, C. J. 2015. *The sexual politics of meat: A feminist-vegetarian critical theory*, Bloomsbury, New York.
- Agamben, G. 2002. *L'aperto. L'uomo e l'animale*, Bollati Boringhieri, Turin; en. tr. *The Open: Man and Animal*, Stanford University Press, Stanford 2004.
- Asafu-Adjaye, J., et al. 2015. *An Ecomodernist Manifesto*, published online at ecomodernism.com
- Bar-On Y. M., Phillips, R., Milo, R. 2018. "The biomass distribution on Earth", *Proceedings of the National Academy of Sciences*, 115(25), pp. 6506-6511.
- Broglio, R. 2013. "When animals and technology are beyond human grasping", *Angelaki*, 18(1), pp. 1-9.
- Burgat, F. 2022. "Will Symbolic Sacrifice Triumph Over Real Sacrifice? A Structuralist Hypothesis on the Role of Meat and Milk Substitutes", en. tr. by Lyman E., James H., *Philosophy Today*, 66(3), pp. 455-469.
- Carary, A., Gruen, L. 2022. *Animal Crisis. A new critical theory*, Polity Press, Cambridge.
- Colombo, R. 2017. "Il dominio sul dissimile", in G. Mormino, R. Colombo, B. Piazzesi, *Dalla predazione al dominio: la guerra contro gli animali*, Raffaello Cortina Editore, Milan.
- Derrida, J. 1991. "Eating Well", or the Calculation of the Subject: An Interview with Jacques Derrida" in E. Cadava, P. Connor, J.-L. Nancy, (eds), *Who Comes After the Subject?*, Routledge, London.
- Fabian E. 9 October 2023. "Defense minister announces 'complete siege' of Gaza: No power, food or fuel", *The Times of Israel*.
- Gaard G. 2012. "Speaking of animal bodies", *Hypatia* 27(2), pp. 520-526.
- Harari, Y. N. 2014. *Sapiens: A brief history of humankind*, Random House, New York.
- Hobbes, T. 1983. *De Cive*, ed. H. Warrender, Clarendon, Oxford.
- Hobbes, T. 2012. *Leviathan*, ed. N. Malcom, Clarendon, Oxford.

- Krzykawski, M. 2021. "Re-animalizing Animals, Re-animating Humans", in C. Mengozzi, (ed.), *Outside the Anthropological Machine*, Routledge, New York.
- Laclau, E. 2003. "Why do empty signifiers matter to politics?", in S. Žižek (ed.), *Jacques Lacan. Critical Evaluations in Cultural Theory*, Routledge, New York.
- Larsen, J., Roney, M. 2013. "Farmed Fish Production Overtakes Beef", *Earth Policy Institute*, http://www.earth-policy.org/plan_b_updates/2013/update114.
- Latour, B. 1993. *We Have Never Been Modern*, Harvard University Press.
- Mengozzi, C. 2021. "Introduction: Ways Out of the Anthropological Machine, or How and Why to Defamiliarize Ourselves", in Idem (ed.), *Outside the Anthropological Machine*, Routledge, New York.
- Mormino G. 2017. "Il sacrificio animale", in G. Mormino, R. Colombo, B. Piazzesi, *Dalla predazione al dominio: la guerra contro gli animali*, Raffaello Cortina Editore, Milan.
- Mormino, G., Colombo, R., Piazzesi B. 2017. *Dalla predazione al dominio: la guerra contro gli animali*, Raffaello Cortina Editore, Milan.
- Patterson N., et al. 2006. "Genetic evidence for complex speciation of humans and chimpanzees", *Nature*, 441, pp. 1103-1108.
- Pellegrino, G., di Paola, M. 2018. *Nell'Antropocene. Etica e politica alla fine di un mondo*, DeriveApprodi, Rome.
- Piazzesi, B. 2017. "La conoscenza degli animali", in G. Mormino, R. Colombo, B. Piazzesi, *Dalla predazione al dominio: la guerra contro gli animali*, Raffaello Cortina Editore, Milan.
- Pimm, S. L., et al. 2014. "The biodiversity of species and their rates of extinction, distribution, and protection", *Science*, 344(6187).
- Pollo, S. 2017. "Going Veg: Care of the Animals or Care of the Self?", *ZOOPHILOLOGICA. Polish Journal of Animal Studies*, 3, 143-154.
- Pollo, S. 2021. *Manifesto per un animalismo democratico*, Carocci, Rome.
- Roser, M. 2023. "How many animals get slaughtered every day?", published online at <https://ourworldindata.org/how-many-animals-get-slaughtered-every-day>.
- Sandom, C., et. al. 2014. "Global late Quaternary megafauna extinctions linked to humans, not climate change", *Proceedings. Biological sciences*, 281(1787).
- Steiner G. 2024. *What We Owe to Nonhuman Animals*, Routledge, New York.

Ivano Zanzarella*

The Problem of Musical Creativity and its Relevance for Ethical and Legal Decisions towards Musical AI

Abstract

Today, AI and technology are no longer simple tools in music production, but active collaborators of human musicians. This raises complex questions about the nature of artificial musical creativity and its ethical and legal implications. This paper addresses such questions by considering a study case, the problem of artificial music authorship. It is argued that attributions of a moral and legal status to Musical AI technologies are closely tied to the theoretical understanding of musical creativity implicitly or explicitly presupposed, and that such attributions are necessary and desirable in present-day society because of pragmatic reasons.

Keywords

Musical AI; Ethics of Artificial Intelligence and Music; Copyright; Computational Creativity; Music Legislation

1. Introduction. Artificial Music: Historical and Conceptual Notes

Making music without musicians has been possible since far before than one may expect (Collins 2018): After some experiments in Classical Antiquity and Middle Ages, the first system for composing music automatically, the *musarithmica mirifica*, was developed by Athanasius Kircher (1650); Games based on dices and cards were introduced in the 18th century for automatizing and randomizing the process of music composition (Gardner 1974; Hedges 1978); Automata and mechanical musical instruments capable of performing music without human intervention were still built during the 20th century (Buchner 1960).

But of course, the field of automatic music composition and performance underwent radical changes with the development of computers

* Università degli Studi di Bari “Aldo Moro”.



and, later, Artificial Intelligence (AI). In 1957, the first musical work entirely composed by a computer, the *Illiac Suite*, was performed (Hiller and Isaacson 1959). Several software have been developed since the '60s with the ability to analyze and encode the style of famous composers and to reproduce it in new compositions – e.g. (Cope 1992). Machine Learning (ML) techniques have made computers able to compose original music even in their own style (Downie 2003; Typke, et al. 2005; Sánchez et al. 2013; Civit et al. 2022). Innovative electronic instruments have been introduced able to synthesize existing and new sounds, making possible to musicians what was unimaginable in the past (Casini and Rocchetti 2018; Miranda 2021).

Yet, while at the beginning algorithmic techniques, computers and AI were nothing but *passive tools* in the hands of composers and performers, today they are becoming *active collaborators* of musicians (Ferry et al. 2019; Miranda 2021). Compositions by computers are regularly performed by prestigious orchestras around the world – that is the case of IAMUS' music performed by the London Symphony Orchestra in 2012 (Sánchez et al. 2013). Music composition and music writing software, sound sampling techniques, audio editing and manipulation software (e.g. Auto-Tune), MIDI sequencers, synthesizers, electronic or digital musical instruments, and various kinds of (AI-powered) music technologies have today taken up a large part of the creative work carried out by musicians in concert halls, stages and recording studios (Katz 2004, 64; Collins 2011, 35; Sturm et al. 2019, 15; Anantrasirichai and Bull 2022).

For this reason, computational musical creativity (CMC) has gained growing attention among AI scientists, musicians and philosophers, with a great number of specialized international conferences and journals dedicating spaces to it and trying to investigate its problems and potential. Besides the theoretical issues concerning CMC, however, scientists and philosophers are reflecting on moral and legal concerns related to the use of artificial music technologies.

The present paper aims to examine the consequences of Musical AI (MAI) on such ethical and legal issues, with a particular focus on the problem of artificial music authorship (Section 2.2.1). The central claim of the paper is that answering questions about the ethics of MAI necessarily requires taking a preliminary stance on the nature of musical creativity in general (Section 2.1). Decisions about considering MAI systems as moral and legal subjects, in fact, depend on the systematic conception of musical creativity one has. In (Section 3) I claim that, from a pragmatic point of view, a computational view of musical creativity should be preferred to a phenomenological one, as this better allows for attributing to MAI technologies a moral and legal status, considered the urgent weight such technologies already have in present-day music industry and society.

2. Theoretical and Ethical Problems of Musical AI

2.1 Philosophy of Music and Musical Creativity

Facing moral and legal issues concerning MAI is impossible without having a previous understanding of musical creativity. Such an understanding can be more easily achieved by considering the following four questions, closely related to each other:

- (Q1) Can an artificial system act creatively as humans do?
- (Q2) Is intentionality a necessary condition for creativity?
- (Q3) Is there something to understand in music?
- (Q4) What does it mean for a machine to be a musician?

Often, creativity is conceived as a “wonder” (Gardner 1974, 135), something which cannot be explained, let alone reproduced artificially (Lovelace 1843). Against this phenomenological conception of creativity, which sees it as something subjective, private, “mystic” and exclusively human, cognitive science – since e.g. (Turing 1950, 450) – has tried to analyze it in more objective terms. In this respect, the ideas of Margaret Boden (2004) have proven to be particularly fruitful.

As a first thing, explaining the “mystery” of creativity does not reduce its significance. Yet, verbal (phenomenological) discourses around it do not add very much to our understanding of the phenomenon, and remain vague about it. For this reason a different approach should be preferred, namely a functionalist and computational one (Boden 2004, 283-284).

According to this view, creative processes can be defined as processes bringing about “products – ideas or artifacts – that are new, surprising, and of value” (Boden 2004, 1). There are three ways (often combined) in which this can be accomplished: by combining familiar ideas (*combinatorial creativity*); searching for new solutions and ideas within a *conceptual space* – a finite system of rules and propositions (*exploratory creativity*); or transforming conceptual spaces themselves (*transformational creativity*). These are the mechanisms underlying creativity¹, and increasing our understanding of creativity amounts to constructing ever more accurate computational models of how they work. Being such, nothing theoretically prevents these mechanisms from being implemented in hardware different from human brains: Creativity emerges merely from the ability of combining, exploring or transforming systems of rules, symbols, etc.

¹ These mechanisms have been also formalized (Wiggins 2006a 2006b; Schmidhuber 2010; Mogensen 2018; Franceschelli and Musolesi 2024).

and is not the result of a some (subjective, phenomenological or divine) inspiration only humans are capable of.

Accepting a computational account of creativity thus answers positively (Q1). However, attributing a quality of an output (creativity) to what produces it (machine), in a sort of behaviorism, may be problematic. The fact that machines can produce something we could regard as creative, in fact, is not enough: Something more is needed in order to consider machines as creative subjects. After all, a creative output may be the result of some stochastic process working in the machine, which would make it creative only in an accidental sense (Casini and Rocchetti 2018, 128-129).

What distinguishes an *accidentally* creative subject, or one which just *acts* creatively, from a truly creative one is namely *intentionality*. Searle (1983, 1; 1998, 85) defines intentionality as the property of a mind to be “directed at, or about, or of, objects and states of affairs in the world” – in other words, the property to have semantic content. In this sense, a subject – human or not – can be said to be (truly) creative if the creative outputs it produces amount to an experienceable representation of the *content* of its mental states – whatever having a mind means for this subject. If so, creativity is not only a formal process in which knowledge is syntactically combined or transformed (a necessary, but not sufficient definition of creativity), but also a process in which, *necessarily* (Q2), formal combinations show semantics and intentionality, acknowledged as such by other subjects (provided with intentionality too)².

Of course, whether machines have (intentional) mental states, and can therefore not only *act* creatively, but *be truly* creative is still not clear (Searle 1980; Boden 2004, 7, 21, 277-304; Gräbe and Kleeman 2022). Thus, the attribution of creativity to machines can occur just in a weak sense. However, if this claim is generally true for the most creative activities, it faces some problems in the case of music.

In fact, according to one of the most widespread philosophies of music, *musical formalism*, music is devoid of meanings in the common sense. It cannot represent external “objects and states of affairs in the world”, and its meaning coincides with its formal structure (Hanslick 1854; Helmholtz 1863; Scruton 1976; Alpersen 2004). In music, meanings cannot be represented in the sense that a particular external reference or sense can be matched to each particular syntactic musical symbol. If music represents something, this is only its syntax (harmonies, chords, scales, etc.).

According to this view, hence, if there is something to understand in music, this is not, in traditional sense, a semantic, representational meaning beyond its syntax, but only the functional and causal role of every

² Further to this, Casini and Rocchetti (2018, 128-129) claim that the creative output should be also the result of a deliberate act of will, which is however controversial.

syntactic element within the specific musical system, form or genre considered (Q3). As a consequence of this, intentionality cannot be considered a necessary condition for musical creativity in particular (Wimsatt and Beardsley 1946; Zimmerman 1966; Carroll 1999; Soldier 2002; Ariza 2009). In fact, if semantics is excluded from music – if, in other words, music cannot express meanings and be “directed at, or about, or of, objects and states of affairs in the world” – creative musical subjects (composers or performers) cannot put in musical outputs the experienceable expression of the intentional content of their ideas and mental states³. Musical creativity, as a result, just derives from a combinatorial game on syntactic musical forms and structures. This game is then supposed to bring about some products to which creativity can be attributed – something carried out generally by listeners, which give music external (mostly emotional) meanings.

The abstractness, self-referentiality and “meaninglessness” of music and musical creativity lead to important consequences for (Q4) and, as it will be noticed, to the ethical (and legal) discourse around MAI. If intentionality is not a necessary condition for musical creativity, in fact, there is no problem in accepting that artificial systems too, which, as mentioned, are supposedly able to *act* creatively, could produce musical outputs that can be regarded as creative. In this sense, they would be creative in music in the same way humans are: If musical creativity develops as a process of syntactic manipulation of musical symbols – which cannot be otherwise, since music has no intentional content – and if artificial systems can act creatively exactly by manipulating syntax, it follows that the latter can potentially⁴ produce creative music as humans are capable to do.

2.2 The Ethics of Musical AI

Today, musicians are delegating increasing portions of music creative processes to MAI (Section 1). Nevertheless, a difference between human and artificial music is still drawn, and resistance exists to accept MAI technologies as autonomous creative subjects – despite their substantial involvement in music creative processes. Just a philosophical justification of the comparability of humans and computers in music creativity on the technical level (Section 2.1) cannot suffice. Something

³ By “ideas”, I mean here the *non-musical ideas*. These latter are ideas which do not regard the syntactic dimension of music and lie beyond it, expressing a sort representational meaning which has to do with things in the world. On the contrary, the only expressible ideas in music can be related to its syntax (forms, structures, musical figures, tones, etc.), i.e. the *musical ideas* (Hanslick 1854; Scruton 1976).

⁴ This means, once they would reach the sufficient level of technical complexity and “manipulative power”.

more is needed for making computers really acceptable as autonomous creative subjects in the field of music, and this is namely a specific *moral (and political) decision of individuals and institutions favoring this acceptance* (Boden 2004, 21), an acceptance indeed required on the ground of the weight MAI technologies already have in current music industry (Section 1; Section 3).

Today, however, even if machines are able to produce music that is “creative” in a very human sense (Section 1; Section 1.2), decisions are rather taken in the opposite direction in order to avoid puzzling moral and legal situations. Yet, are moral and legal decisions depriving machines of an autonomous creative status really necessary and desirable in our society? In the following, I shall investigate this question on the basis of a case study: the problem artificial music authorship.

2.2.1 The Problem of Artificial Music Authorship

The authorship of artificially composed or performed music is one of the most urgent ethical and legal issues deriving from the use MAI technologies.

In 2010, a group of singers (Hatsune Miku, Kagamine Rin and Len, and Utatane Pik) reached the top of music charts and achieved great sales reports in Japan. Interestingly, the group was composed by virtual human-like characters, who performed music entirely composed by a software, Vocaloid, developed and distributed by the Japanese corporation Yamaha. Vocaloid is a singing voice synthesizer software that allows to synthesize songs for fictional singers virtually embodied by anime and manga avatars (just like Hatsune Miku and her colleagues). The original voices implemented in the software, which can be then modified, altered, adapted to certain lyrics and melodies etc. derive from the sampled voice of real human singers, and the fictional avatars are also created by human designers.

This means, a musical product created by means of Vocaloid has a multiple authorship since the subjects concurring in the creation process are many: the singer whose voice was sampled, the designer of the avatars, the programmer of the software, the company which distributes it, the “composer” who sets up, triggers and uses the program she paid for, expressing in this way her musical ideas, and possibly, if different from that composer herself, the author of the lyrics and the melody (or, still, the programmer who produced some algorithms and software for producing them). For this reason, apportioning appropriately credits, authorship and rights to the final product of Vocaloid between these subjects is something extremely challenging, especially for institutions and lawgivers (Collins 2011, 36).

The same condition applies, for example, to another automatic music generation software, FolkRNN, an open source software available online (<https://folkrrnn.org/>) which uses recurrent neural networks and AI for analyzing great quantities of folk music transcriptions in internet and producing new folk tunes from that.

Regulating the issue of the authorship of the music produced by software such as those mentioned above is a task that has been undertaken by several countries around the world (Sturm et al. 2019). On the one hand, countries like UK, Hong Kong, South Africa, India, Ireland and New Zealand adopted copyright laws envisaging the operator of the given computer program as the owner of the rights to the musical works generated by it (Sturm et al. 2019, 4). This is evidently an effect of a moral and political decision which does acknowledge an active role within the artificial creative process both to the software – thus not solely to the creators and programmers of it (see EMI’s case below) – *and* to the human subject operating it and making all “the arrangements necessary for the creation of the work” (Sturm et al. 2019, 4). This means that the artificial system is generally considered a creative co-author of the work, whilst, however, it is always the human subject to bear entirely the *legal* responsibility for uses and misuses of its products. That artificial systems for music generation are, in these countries, at the center of a moral and political decision which acknowledges them from the moral point of view as (at least partly) autonomous creative subjects appears clearer and even more emphasized particularly in the British copyright law, that defines computer-generated works as works “generated by computer in circumstances such that *there is no human author of the work*” (Copyright, Designs and Patents Act 1988, I(X), S. 178, my italics). For what concerns the legal point of view, such legislations do not allow however to regulate the matter in a homogeneous, univocal and unambiguous way and are therefore insufficient (Sturm et al. 2019, 4). For example, it remains unclear what the precise legal responsibilities of the programmer and the user are, or in which terms the software itself could be considered legally relevant.

On the other hand, there are countries like the USA or most of Europe, in which an opposite moral and legal decision is taken towards software of automatic music generation (Sturm et al. 2019). In these countries artificial systems capable of creative products certainly gain a form of aesthetic acknowledgment, which does not however have moral and legal effects: Creativity is a category belonging only to humans and as such it has to be understood and treated by institutions and legal systems. If any at all, provisions regulating the authorship of computer-generated music basically apply the same legal categories of “normal” copyright laws resting on “human-centered concepts, both with regards to the beneficiary

of protection (i.e. the author), the conditions for protection (e.g. originality), and the rights granted (economic, but also moral rights)” (Sturm et al. 2019, 4). The Court of Justice of the European Union (CJEU), for example, considers a work original or creative if it is expression of the author’s free creative choices, personality, or personal touch⁵ (Sturm et al. 2019, 4). In quite the same terms, taking another example, no credits are assigned to the program EMI itself by his creator, the American scientist and composer David Cope, who conversely owes all rights to the works generated by it: The program is considered incapable of autonomous creativity, “the hand of the composer is not absent from (its) finished product” (Cope 1991, 2).

In this way, however, every possibility to morally acknowledge creativity in machines is excluded a priori and, with it, every possibility to have a copyright law protecting machine authorship and AI-generated works (Guadamuz 2017; Ramalho 2017; Buning 2018; Michaux 2018; Deltorn and Macrez 2019; Lauber-Rönsberg and Hetmank 2019; Sturm et al. 2019, 4). Yet, on the basis of the theoretical premises from (Section 2.1), it is simple to see that this particular moral and thus legal *decision* regarding AI-systems involved in creative processes as incapable of autonomous creativity and originality eventually depends on a particular, *implicit* account of creativity which makes this latter dependent from phenomenological and subjective factors (the author’s personal experiences, touch, will etc.). As we observed, this is, after all, only *one* account of creativity, and maybe also not the more accurate one, at least in the case of music. Especially with (Boden 2004), in fact, it has been pointed out in (Section 2.1) that such verbal theories of creativity are too naïve, and therefore insufficient, for giving a realistic account of how creative processes really work and what the necessary conditions for creativity – and musical creativity in particular – really are. Certainly – in the absence of an AGI – phenomenological factors are still important, for instance in triggering creative processes, but, at least for what concerns musical creativity from a technical point of view, they appear not to be so indispensable for the *intrinsic* functioning of musical creative processes, which have rather to do with syntax than semantics and phenomenology (Section 2.1; Ariza 2009, 65). Moreover, originality too results a vague notion in music. This concept cannot be simply taken from the everyday language and used, maybe in judgments and laws, without defining critically in advance a univocal and appropriate meaning of it. In which sense, for example, is

⁵ Infopaq: C-5/08, Judgment of the Court (Fourth Chamber) of 16 July 2009; BSA: C-393/09, Judgment of the Court (Third Chamber) of 22 December 2010; Painer: C-145/10, Judgment of the Court (Third Chamber) of 1 December 2011; Dataco: C-604/10, Judgment of the Court (Third Chamber) of 1 March 2012.

Mozart's music original compared to Bach's, considering that the former spent years studying and assimilating the contrapuntal innovations of the latter? What piece of music, composed by humans, could be truly defined as something *completely* original and (syntactically⁶) unrelated with the rest of the contemporary and previous music compositions? (Casini and Rocchetti 2018, 129).

It may be conceived, for example, of an AI-powered automatic music generation software which is able to compose some pieces of music in Mozart's style as a human composer too would also be capable to do. We admit that both internalized every composition of the Austrian composer (of course in qualitatively different ways). The software, due to its higher computational skills, could explore even more (and therefore more original) musical combinations formally coherent with Mozart's style than a human would do. Now, regardless of the aesthetic significance of the outcomes of both, in what would the compositions by the software be less or more original than those by the human? And, in general, why is it (arbitrarily) presupposed that (subjective) aesthetic criteria can be used for judging the intrinsic originality (or creativity) of a musical piece? Does this not depend, *in music*, only on a matter of syntactic combinations? (Section 2.1). Furthermore, the (more elaborate) combinations found by the software could be, at the end, only indirectly and indeed very hardly brought back to the actual intentions, experiences, touch, will, etc. of its human programmer: They are actually so far, unimaginable and unpredictable for her that no direct "phenomenological" interdependence connection can be truly stated between the two. Yet, CJEU's judgments, for example, require such a connection as a necessary condition for the attribution to the latter of moral and legal rights to the authorship of the former. Even if the existence of this connection cannot be really detected, the attribution most of the times takes place anyway: This clearly shows a theoretical bias in such kind of legislations.

Thus, copyright laws which take the originality of a musical work as a necessary condition for the attribution of moral and legal rights to its creator are simply considering the problem from a perspective which is too narrow and not critical enough. Originality cannot be said to be a necessary condition for musical creativity, but only a sufficient one (Section 2.1), and a phenomenological account of musical creativity cannot serve as a basis for the attribution (or in this case non-attribution) of moral and legal rights to MAI technologies⁷.

⁶ Semantic relations between musical pieces are after all practically impossible to detect, since, as we know, music is not representational.

⁷ I want to stress that this conclusion is valid only for what concerns music and MAI and not for other fields in artificial creativity.

It appears however reasonable that a more accurate and comprehensive legislation about machine authorship should be developed. Automatic systems for music composition and performance are becoming ever more autonomous, i.e. capable of truly original and creative products (Collins 2011, 37-38; Casini and Rocchetti 2018, 127; Sturm et al. 2019, 4; Section 2.1), and, indeed, they already play a significant role in present-day music industry and society (Section 1). This means, consequently, that some kind of moral and legal responsibility should be acknowledged to them. However, finding a way in which this could concretely happen currently represents one of the most important and difficult challenges for many countries around the world, including EU (Sturm et al. 2019, 4). The possible perspectives in this sense are different. Perhaps “authorship recognition may require an analysis of the operation of the systems and the role of the different actors involved in the process (e.g., the developer, the trainer or the user)” (Sturm et al. 2019, 4). If MAI-systems will be able to replicate famous composers or even performers, changes in the definition itself of copyright and authorship may also be required: “[C]opyright will be perpetual [...] or effectively lawless or, most likely, will remain somewhere complexly in-between” (Collins 2011, 38), “adjustments may be needed to the existing framework to either amend the existing copyright laws or to pass new *sui generis* rights targeting AI-generated products” (Sturm et al. 2019, 5, italics in original) – see also (Schafer et al. 2015). The owner of the copyrights on AI-generated music could no longer be determined by traditional categories like intellectual property, expression of personal creative choices, ideas, will, etc. but merely through economical ones – simply, who paid for the program has copyrights on its products (Sturm et al. 2019, 9). Or still, “machines”, when powerful and intelligent enough in a probable future, “may at some point stand up for their own IP⁸ rights as dynamic creators, whilst the existing big content companies will fight to retain power as long as they can by denying that AIs have reached sufficient independence” (Collins 2011, 38).

3. Conclusion

In the previous section I examined one of the most relevant and urgent ethical problems of MAI, namely artificial music authorship, considering advantages and shortcomings of the different ways in which countries legally regulate the attribution of authorship to artificial systems, as well as some possible future perspectives on the topic.

⁸ IP: Intellectual Property.

Here, it has been noticed that beyond acknowledging the creativity of MAI technologies *at a mere aesthetic level* – this is eventually the reason why they are used by musicians – a *moral role* is explicitly or implicitly attributed to them, which is necessary because of the social relevance they already gained (they can intervene in music industry and interact with human musicians in working relationships, they are legally relevant subjects in the matter of the attribution of copyrights, etc.). A moral decision about MAI technologies is also necessary as a basis for issuing legal provisions defining their social role and regulating the relation between them and humans. This appears particularly clear with regard to the problem of machine musical authorship. Moreover, it has been pointed out that the kind of moral decision taken, i.e. the way in which the moral and social role of such artificial systems is acknowledged, essentially determines the legal provisions issued.

As emerged from the discussion of CJEU's judgments and of the problem of originality in (Section 2.2.1), every moral decision towards artificial systems of music production strictly depends on the theoretical account of music creativity explicitly or implicitly adopted. More in general, the way – and indeed the possibility itself – of solving the ethical problems of Musical AI depends on what is considered as characteristic of musical creativity. By accepting a formalistic, computational or syntactic theory of musical creativity, which explains human and artificial musical creative processes in the same computational terms (Section 2.1), the possibility is admitted to consider human and artificial creative subjects in music morally in the same way and at the same level. Conversely, defending a theory which takes phenomenological and aesthetic factors as necessary conditions for musical creativity (experiences, will, touch, etc. of the author, intrinsic aesthetic value of the musical work) leads to the impossibility of acknowledging artificial systems of music production (still in absence of a human-like AGI) as autonomous moral subjects.

Certainly, the adoption of a given theoretical account on musical creativity, and thus the moral decisions which derive from that, also depend on the particular technology considered, and on the specific ethical and social context in which such decisions have to be taken. This is the reason why these two perspectives on (artificial) musical creativity – the computational and phenomenological one – still remain not mutually exclusive. Beyond the issue of the theoretical and scientific suitability of each of them, however, from a merely pragmatic point of view, the computational perspective actually seems to be the most helpful one: Since artificial systems of music production are *already* part of the everyday life and everyday work of musicians, music industry operators and music consumers (Section 1), we *already* need to attribute to them a moral status, after which, by the way, a corresponding definition of their legal condition,

and a corresponding formulation of laws and provisions on their use, can only become possible.

It is inevitable, in general, that we will become more and more technology- and AI-entangled in the future, that technology and AI will gain ever larger importance in every aspect of our life. Accordingly, artificial systems of music production too will be considered as ever more autonomous agents, and AI-generated music as music in its own right. If so, the need of learning, in some way, to morally, socially and legally live and cope with this emerges with urgency (Collins 2011; Casini and Rocchetti 2018; Sturm et al. 2019).

For this reason, moral decisions favoring the recognition of MAI systems as (moral) creative subjects, is today highly desirable. Nevertheless, it is necessary that the theoretical presuppositions on the basis of which such decisions are taken are always made explicit. Furthermore, no decision should be taken before discussing and defining sharply its alleged theoretical presuppositions. From this point of view, the importance of philosophy and science is obvious: Philosophy and aesthetics of music, philosophy of mind and AI, computer sciences, cognitive sciences, neurosciences, etc. help us a lot in understanding how music creative processes work, what musical creativity is, what we should regard as truly creative in music, etc. and, therefore, to appropriately attribute a moral and legal status to MAI technologies.

References

- Alperson, P.
2004 *The Philosophy of Music: Formalism and Beyond*, in Kivy, P. (ed.), *The Blackwell Guide to Aesthetics*. Blackwell, Hoboken.
- Anantrasirichai, N., Bull, D.
2022 *Artificial Intelligence in the Creative Industries: A Review*, in “Artificial Intelligence Review”, 55(1), 589-656.
- Ariza, C.
2009 *The Interrogator as Critic: The Turing Test and the Evaluation of Generative Music Systems*, in “Computer Music Journal”, 33(2), 48-70.
- Boden, M.
2004 *The Creative Mind: Myths and Mechanisms*, Psychology Press, Hove (UK).
- Buchner, A.
1960 *Mechanical Musical Instruments*, Batchworth Press, Buckland.
- Buning, M.C.
2018 *Artificial Intelligence and the Creative Industry: New Challenges for the EU Paradigm for Art and Technology*, in Barfield, W. and Pagallo, U. (eds), *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar Publishing, Cheltenham.

- Carroll, N.
1999 *Philosophy of Art: A Contemporary Introduction*, Routledge, London.
- Casini, L., Rocchetti, M.
2018 *The Impact of AI on the Musical World: Will Musicians be Obsolete?*, “Studi di Estetica”, IV.
- Civit, M., Civit-Masot, J., Cuadrado, F., Escalona, M. J.
2022 *A Systematic Review of Artificial Intelligence-Based Music Generation: Scope, Applications, and Future Trends*, “Expert Systems with Applications”, 209, 118-190.
- Collins, N.
2011 *Trading Faures: Virtual Musicians and Machine Ethics*, “Leonardo Music Journal”, 21, 35-39.
- 2018 *Origins of Algorithmic Thinking in Music*, in McLean, A., Dean, R. T. (eds), *The Oxford Handbook of Algorithmic Music*, Oxford University Press.
- Cope, D.
1991 *Computers and Musical Style*, Oxford University Press.
1992 *Computer Modeling of Musical Intelligence in EMI*, “Computer Music Journal”, 16(2), 69-83.
- Deltorn, J.M., Franck, M.
2019 *Authorship in the Age of Machine learning and Artificial Intelligence*, in *The Oxford Handbook of Music Law and Policy*, Oxford University Press.
- Downie, J.
2003 *Music Information Retrieval*, “Annual Review of Information Science and Technology”, 37, 295-340.
- Ferry, M., Surya, C., Hidayat, Y., Pemady, M.P.
2019 *The Value of Musical Creativity in Industry 4.0 Era: Based on Musical Composition Generated by Artificial Intelligence & Computer Learning*, in *Proc. of the 1st International Conference on Intermedia Arts and Creative Technology*, Yogyakarta, Indonesia, 185-191.
- Franceschelli, G., Musolesi, M.
2024 *Creativity and Machine Learning: A Survey*, “ACM Computing Surveys”, 56(11), 1-41.
- Gardner, M.
1974 *Mathematical Games*, “Scientific American”, 231(6), 132-137.
- Gräbe, H.G., Kleemann, K.P.
2023 *Intentionality and Machines*, “Philosophia”, 51(2), 675-695.
- Guadamuz, A.
2017 *Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works*, “Intellectual Property Quarterly”, 2, 169-186.
- Hanslick, E.
1854 *Vom musikalisch Schönen*, Weigel, Leipzig.
- Hedges, S.A.
1978 *Dice Music in the Eighteenth Century*, “Music & Letters”, 59(2), 180-187.
- Helmholtz, H.
1863 *Die Lehre Von Den Tonempfindungen*, Vieweg, Braunschweig.

- Hiller, L., Isaacson, L.
1959 *Experimental Music: Composition with an Electronic Computer*, McGraw-Hill, New York.
- Katz, M.
2004 *Capturing Sound: How Technology Has Changed Music*, California University Press, Berkeley.
- Lauber-Rönsberg, A., Hetmank, S.
2019 The Concept of Authorship and Inventorship under Pressure: Does Artificial Intelligence Shift Paradigms? "Journal Intellectual Property Law & Practice", 14, 570-579.
- Lovelace, A.K.,
1843 *Sketch of the Analytical Engine Invented by Charles Babbage [...] With Notes by the Translator*, "Scientific Memoirs", 3(29), 666-731.
- Michaux, B.
2018 *Singularité technologique, singularité humaine et droit d'auteur*, in Poullet, Y. (ed) *Droits, Norms et Libertés Dans le Cybermonde*, Larquier, Paris, 401-416.
- Miller, A.
2019 *The Artist in the Machine: The World of AI-Powered Creativity*, MIT Press, Cambridge (MA).
- Miranda, E.R.
2021 *Handbook of Artificial Intelligence for Music*, Springer, New York.
- Mogensen, R.
2018 *Dynamic Concept Spaces in Computational Creativity for Music*, "Philosophy and Theory of Artificial Intelligence" 2017, 57-68.
- Sánchez, Q.
2013 *Melomics: A Case-Study of AI in Spain*, "AI Magazine", 34(3), 99-103.
- Schafer, B.
2015 *A Fourth Law of Robotics? Copyright and the Law and Ethics of Machine Co-Production*, "Artif. Intell. Law", 23, 217-240.
- Schmidhuber, J.
2010 *Formal Theory of Creativity, Fun, and Intrinsic Motivation*, "IEEE Transactions on Autonomous Mental Development", 2(3), 230-247.
- Scruton, R.
1976 *Representation in Music*, "Philosophy", 51(197), 273-287.
- Searle, J.
1983 *Intentionality. An Essay in The Philosophy of Mind*, Cambridge University Press, Cambridge.
- Searle, J.
1998 *Mind, Language and Society: Philosophy In The Real World*, Basic Books, New York.
- Soldier, D.
2002 *Eine Kleine Naughtmusik: How Nefarious Nonartists Cleverly Imitate Music*, "Leonardo Music Journal", 12, 53-58.
- Sturm, B.L.T.
2018 *Artificial Intelligence and Music: Open Questions of Copyright Law and Engineering Praxis*, "Arts", 8(3), 115.

Turing, A.

1950 *Computing Machinery and Intelligence*, "Mind", 59(236), 433.

Typke, R., Wierig, F., Veltkamp, R.

2005 *A Survey of Music Information Retrieval Systems*, in *Proceedings of the 6th International Conference on Music Information Retrieval*. Queen Mary.

Wiggins, G.A.

2006a *A Preliminary Framework for Description, Analysis and Comparison of Creative Systems*, in "Knowledge-Based Systems", 19(7), 449-458.

2006b *Searching for Computational Creativity*, "New Generation Computing", 24, 209-222.

Wimsatt, W.K., Beardsley, M. C.

1946 *The Intentional Fallacy*, *Sewanee Review*, 54, 468-488.

Zimmerman, R.L.

1966 *Can Anything Be an Aesthetic Object*, "The Journal of Aesthetics and Art Criticism", 25(21), 177-186.

Discussioni

Intelligenza Artificiale: sfide e controversie.

A cura di Maria Filomena Anzalone e Maria Russo

*Giuseppe De Ruvo**

Intelligenza artificiale, sicurezza nazionale e capitalismo politico. La crisi della riflessività democratica e la necessità di un'etica strategica

Abstract

Artificial intelligence, national security and political capitalism. The crisis of democratic reflexivity and the need for strategic ethics

Focusing on the U.S. case and drawing on the notion of political capitalism, the article thematizes the relationship between AI governance and the concept of national security, often invoked to protect the American tech ecosystem from China's technological rise. We will therefore analyze the relationship between national security apparatuses, the market, and public deliberative practices. We argue that the difficulties in implementing AI governance depend not only on the reluctance of Big Tech, but also – and more importantly – on their ties to defense bureaucracies. The article claims that a critical theory of AI cannot be limited to diagnosing and mitigating particular problems emerging from its implementation. Rather, it must deeply thematize the changing relationship between the state and the market, also considering the geopolitical and geoeconomic issues that characterize the political economy of AI. In conclusion, we argue that, for Western countries, taking into account the ethical issues posed by new technologies may also be considered a strategic operation.

Keywords

Artificial Intelligence; Critical Theory; Geopolitics; National Security; Political Capitalism

1. L'irrefrenabile ascesa dell'intelligenza artificiale e della sicurezza nazionale

Il rapido e dirompente sviluppo delle tecnologie legate all'intelligenza artificiale non è neutro. Al contrario, date le sterminate possibilità di applicazione in ambito geopolitico-militare – sia a livello di intelligenza sia

* Università Vita-Salute San Raffaele.



di effettivo combattimento¹ – esse vengono sempre più spesso considerate come degli asset critici per il mantenimento della sicurezza nazionale. Questo concetto, a sua volta, è in rapida espansione. La *national security* non riguarda più solamente le minacce al territorio di un paese, ma arriva a integrare anche la necessità di mantenere posizioni di *leadership* nello sviluppo di tecnologie strategiche e nelle catene del valore legate a queste ultime². In un siffatto contesto, da un lato lo Stato tutela le aziende strategiche dalla concorrenza di paesi considerati rivali, bloccando tentativi di acquisizione da parte di attori (privati o non) riconducibili a essi³ o imponendo sanzioni⁴; dall'altro, esso favorisce la filiera produttiva dei propri “campioni nazionali”, sia attraverso sussidi sia mediante la creazione di ecosistemi favorevoli al loro operato da un punto di vista legale ed economico⁵.

Si inaugura dunque un meccanismo di *partnership* pubblico-privato irriducibile alla mera acquisizione da parte dello Stato di quote azionarie delle aziende. Piuttosto, in quello che la letteratura oramai definisce “capitalismo politico”, assistiamo:

1. a un'informale alleanza tra grandi conglomerati industriali e agenzie di sicurezza nazionale, caratterizzata da meccanismi di *sliding doors* tra governo e industria e dalla “collaborazione, in vista di benefici comuni, tra *élites* politiche ed economiche”⁶;
2. alla divisione del mercato mondiale in “aziende strategiche” e “aziende nemiche”, le quali vengono rispettivamente “protette o avversate [...] soprattutto tramite direttive politiche, burocrazie statali dedicate, norme speciali”⁷;
3. all'estendersi del concetto di sicurezza nazionale, il ricorso al quale offre allo Stato l'autorità per bloccare operazioni economiche che potrebbero in qualche modo ledere i suoi interessi, indipendentemente dall'antieconomicità della decisione⁸.

¹ Si vedano M. A. Flournoy, *AI is already at war*, “Foreign Affairs”, 6 (2023), pp. 55-70; A.B. Zegart, *Lies, Spies and Algorithms*, Princeton University Press, Princeton 2022.

² Per l'allargamento, anche giuridico, di questa nozione, cfr. L. Picotti, *La legge del più forte. Il diritto come strumento di competizione tra Stati*, Luiss University Press, Roma 2023.

³ Rivelativo è il caso Broadcom/Qualcomm: gli USA hanno bloccato, per motivi di sicurezza nazionale, un affare da 103 miliardi di dollari in virtù di presunti legami (mai dimostrati) tra Broadcom e Pechino. Si veda, A. Aresu, *Le potenze del capitalismo politico. Stati Uniti e Cina*, La Nave di Teseo, Milano 2020, pp. 378-381.

⁴ N. Mulder, *The economic weapon. The rise of sanctions as a tool of modern war*, Yale University Press, New Haven 2023.

⁵ M. Mazzucato, *Lo Stato Innovatore*, Laterza, Bari 2014.

⁶ R. G. Holcombe, *Political Capitalism*, “Cato Journal”, 1(2015), p. 41.

⁷ A. Aresu, *op. cit.*, p. 15.

⁸ J. Benton Heath, *The new security challenge to economic order*, “The Yale Law Journal”, 1(2019), pp. 1048-1049.

In tale congiuntura, lo sviluppo dell'IA non può essere slegato dalle dinamiche geoeconomiche e geopolitiche e dal ruolo preponderante che gli Stati – in particolare USA e Cina – hanno in questo ambito. Tuttavia, il riferimento sempre più frequente a una nozione giuridicamente vaga e discrezionale⁹ come quella di sicurezza nazionale e il potere sempre più evidente delle burocrazie statali nel prendere decisioni strategiche generano dei problemi etico-politici non trascurabili, primo fra tutti quello di una pressoché totale assenza di deliberazione pubblica sulle questioni tecnologiche.

Nel prossimo paragrafo ci concentreremo dunque sulla ricostruzione degli equilibri geopolitici che hanno portato all'ampliamento della nozione di sicurezza nazionale, per poi discutere una possibile via d'uscita dalle strettoie imposte dal nuovo rapporto che si è istituito tra Stato, mercato e burocrazie securitarie. In conclusione, si cercherà di mostrare come lo sviluppo di un'etica critica dell'IA non debba essere inteso in opposizione alla legittima e necessaria tutela della sicurezza nazionale.

2. Il conflitto tecnologico sinoamericano e la risurrezione dell'apparato militare-industriale

In un editoriale apparso sul *New York Times* nel febbraio del 2020, Eric Schmidt – fondatore di Google e oggi direttore della *National Security Commission on AI* (NSCAI) – scrive:

Negli ultimi anni, gli americani [...] si sono affidati troppo al settore privato per garantirsi il primato nelle nuove tecnologie. Oggi siamo in una competizione tecnologica con la Cina [...]. Il governo deve tornare seriamente in campo. Le tendenze non ci sorridono. Il nostro primato nell'IA è precario [...]. Abbiamo bisogno di una partnership mai vista prima tra governo e industria [...]. Se vogliamo la vittoria del modello americano, deve essere il governo a comandare.¹⁰

La (non più) silente ascesa della Cina è la principale causa dell'espansione del concetto di sicurezza nazionale¹¹. Grazie al suo peculiare sistema economico, Pechino riesce a favorire l'innovazione senza rinunciare alla concorrenza, favorendo la creazione di *cluster* d'innovazione da cui

⁹ L. Picotti, *op. cit.*

¹⁰ E. Schmidt, *I used to run Google. Silicon Valley could lose to China*, "The New York Times", 27/2/2020.

¹¹ D. Wang, *China's hidden tech revolution*, "Foreign Affairs", 2 (2023), pp. 64-77. Si veda anche R. Lightizer, *No trade is free. Changing course, taking on China and helping American workers*, Broadside Books, New York 2023. Lightizer è stato rappresentante del commercio degli Stati Uniti dal 2017 al 2021.

emergono campioni nazionali in grado di competere con le aziende occidentali e pronti a seguire le direttive (geo)politiche del Partito Comunista Cinese¹².

Per quanto riguarda l'IA, la Cina – seguendo questo schema – è riuscita in pochi anni a ridurre il *gap* nei confronti dell'America: essa “possiede la forza, il talento e l'ambizione per superare gli Stati Uniti come leader mondiale nell'IA nel prossimo decennio”¹³. Il modello cinese offre infatti a Pechino dei clamorosi vantaggi competitivi, anche grazie a un sistema di sorveglianza governativo estremamente diffuso e integrato con il settore privato. Questa organizzazione, senza dubbio autoritaria, mette a disposizione dell'apparato tecno-economico della Repubblica Popolare un'enorme mole di dati per allenare le intelligenze artificiali¹⁴. Come è stato notato, infatti,

La Cina ha già superato gli Stati Uniti per quanto riguarda la quantità totale di dati che è in grado di trattare e produrre. Questa raccolta dati non è impressionante solo per la quantità, ma grazie allo straordinario ecosistema tecnologico cinese – un universo alternativo di prodotti e funzioni mai visti – questi dati sono fatti su misura per costruire aziende legate all'intelligenza artificiale in grado di fare un grande profitto.¹⁵

Inoltre, nel *New Generation Artificial Intelligence Development Plan 2017-2030*, Pechino specifica chiaramente come lo sviluppo dell'IA, portato avanti in simbiosi con le aziende private, debba avere un'immediata ricaduta militare: “lo sviluppo dell'IA è decisivo strategicamente per rilanciare la competitività nazionale e proteggere la sicurezza nazionale [...] facendo un salto di qualità nelle nostre capacità militari”¹⁶. Fondamentalmente, l'obiettivo di Pechino è quello di sviluppare una forma asimmetrica di “guerra intelligente” (*zhinenghua*) che possa permetterle di sopperire, attraverso l'IA, alla strutturale inferiorità nei confronti delle Forze armate americane¹⁷. L'obiettivo strategico è ristabilire la deterrenza nello Stretto di Taiwan ed essere pronti a un conflitto asimmetrico¹⁸.

¹² Per una ricostruzione dettagliata: M. Wu, *The “China, Inc.” Challenge to Global Trade Governance*, “Harvard International Law Journal”, 2 (2016), pp. 262-324.

¹³ NSCAI, *Intelligenza Artificiale e Sicurezza Nazionale*, in G. De Ruvo (a cura di), *Storia e Filosofia della Geopolitica*, Carocci, Roma 2024, p. 185.

¹⁴ M. Taylor, *China's Digital Authoritarianism*, Palgrave MacMillan, New York 2022.

¹⁵ K-F. Lee, *AI superpowers*, Houghton Mifflin Harcourt, Boston 2018, p. 23.

¹⁶ La traduzione del documento è disponibile qui: <https://tinyurl.com/47wnb9f8>.

¹⁷ Cfr., H. Roberts *et alia*, *The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation*, “Ai & Society”, 1(2021), p. 62.

¹⁸ Cfr., R. Doshi, *The Long Game: China's Grand Strategy to Displace American Order*, Oxford University Press, Oxford 2021.

La reazione americana, seppur tardiva, è stata decisa e tripartita¹⁹. Da un punto di vista “offensivo”, Washington ha sfruttato la sua posizione di vantaggio nelle catene del valore per limitare la capacità di Pechino di produrre *AI chips*, componenti critiche per l’effettiva realizzazione dell’IA da cui la Cina è ancora dipendente²⁰. Da un punto di vista “difensivo”, gli apparati burocratici di sicurezza nazionale americana hanno invece iniziato a scrutinare con maggiore attenzione gli investimenti cinesi nei settori strategici legati all’IA. Il *Committee of Foreign Investments in the US* (CFIUS), istituto che ha il compito di valutare se gli investimenti diretti esteri presentino dei rischi alla sicurezza nazionale, si è trasformato, secondo alcuni studiosi, in un “strumento dell’arte di governo americana che prende di mira la Cina”²¹.

Infine, vi è la questione della riorganizzazione del rapporto tra governo e industria, che in questa sede ci interessa particolarmente per le sue ricadute etico-politiche. La questione centrale posta dalla NSCAI è infatti quella di “coniugare la leadership *top-down* con l’innovazione *bottom-up*”²². Concretamente, ciò significa affrontare un problema fondamentale, così riassunto dal China Strategy Group (CSG): “il governo americano non beneficia dei dati raccolti dal settore privato e il settore privato non beneficia dei dati raccolti dal governo”²³. Ovviamente, ciò non significa permettere al governo di accedere ai dati personali dei cittadini raccolti dalle piattaforme digitali per allenare le IA, come avviene in Cina. Tuttavia, proprio per questa ragione, le operazioni di queste aziende devono essere *tutelate* sia da influenze estere sia dai tentativi di regolamentazione interna: “gli obiettivi di sicurezza del governo americano impongono una stretta collaborazione tra esso e le aziende tecnologiche private”²⁴. Insomma, lo Stato, attraverso meccanismi di *lawfare*²⁵, protegge le aziende dalla concorrenza straniera e le tutela, come la giurisprudenza degli ultimi anni dimostra²⁶, anche da poten-

¹⁹ Per un’analisi estremamente dettagliata dei singoli casi, si veda A. Aresu, *Geopolitica dell’Intelligenza Artificiale*, Feltrinelli, Milano 2024.

²⁰ Cfr., C. Miller, *Chip War*, Garzanti, Milano 2023.

²¹ U. Khanapurkar, *CFIUS 2.0: an instrument of American statecraft targeting China*, “Journal of Current Chinese Affairs”, 2(2020), pp. 226-240.

²² NSCAI, *op. cit.*, p. 206.

²³ CSG, *Competizione Asimmetrica. Una strategia tecnologica contro la Cina*, in G. De Ruvo (a cura di), *op. cit.*, p. 197.

²⁴ A. Bradford, *Digital Empires*, Oxford University Press, Oxford 2023, p. 58.

²⁵ O.F. Kittrie, *Lawfare*, Oxford University Press, Oxford 2016. Si veda anche, per una prospettiva che tematizza il rapporto tra diritto e geopolitica da un punto di vista più concettuale e non limitatamente al caso americano, N. Irti, *Norma e Luoghi. Problemi di geodiritto*, Laterza, Roma-Bari 2001.

²⁶ Su come le Corti americane abbiano storicamente tutelato le aziende digitali, cfr., A. Bradford, *op. cit.*, pp. 46-47.

ziali regolamentazioni interne, che potrebbero minarne il potenziale d'innovazione, considerato strategico ed essenziale per il mantenimento della sicurezza nazionale.

Il capitalismo politico americano si basa dunque sull'interazione tra potere esecutivo, giudiziario ed economico. Per quanto, nelle questioni attinenti alla sicurezza nazionale, l'ultima parola spetti di norma al presidente, nondimeno vi è da sottolineare come la maggior parte del lavoro – nonché l'individuazione delle minacce – sia svolto dagli apparati di sicurezza (come il CFIUS). Allo stesso tempo, i report di pianificazione strategica – da cui emergono questioni strettamente collegate alla governance dell'IA, anche in ambito civile – vengono elaborati da commissioni (come la NSCAI o il CSG) prive di qualsiasi legittimità democratica. In un siffatto processo, il presidente, formalmente garante dell'unità nazionale e unica figura dotata di legittimità popolare, assomiglia al sovrano descritto da Hegel nell'appendice al §280 dei *Lineamenti di filosofia del diritto*: “in una organizzazione compiuta è soltanto il culmine del decidere formale a importare, e per monarca occorre soltanto un uomo che dice ‘sì’ e mette il punto sulla *i*”²⁷.

Questo approccio sta permettendo agli USA di mantenere il primato tecnologico nei confronti della Cina. L'enorme apparato burocratico americano sta escludendo Pechino dalle catene del valore necessarie per sviluppare IA di alto livello e sta tutelando le imprese nazionali, le quali paiono aver recepito le indicazioni (e soprattutto i sussidi) del governo²⁸. I rapporti sempre più stretti tra Big Tech e apparati di sicurezza stanno infatti permettendo alle prime di continuare a operare in maniera piuttosto libera, senza doversi scontrare con un impianto normativo stringente come quello europeo. In cambio, le grandi aziende tecnologiche – oltre che migliaia di start-up – si impegnano a mettersi a disposizione dello Stato, collaborando con gli apparati di sicurezza nazionale e sviluppando sinergie con le burocrazie strategiche:

Le aziende tecnologiche si sono mostrate desiderose di lavorare con il dipartimento della Difesa sull'IA. Quando il Pentagono ha chiesto offerte per un contratto di cloud computing da 10 miliardi di dollari basato sull'IA, Amazon, IBM, Microsoft e Oracle hanno combattuto per aggiudicarselo. Inoltre, va rilevata un'esplosione di start-up che si occupano di sistemi di IA orientati alla difesa.²⁹

²⁷ G.W.F. Hegel, *Lineamenti di filosofia del diritto*, Laterza, Roma-Bari 2019, p. 371.

²⁸ Il solo *CHIPS and Science Act* garantisce 50 miliardi di dollari per l'innovazione tecnologica, a cui vanno aggiunti i denari mobilitati dal *Chips Act* e dall'*Inflation Reduction Act*. In totale, tra 2022 e 2024, la somma mobilitata si aggira intorno ai 300 miliardi di dollari.

²⁹ P. Scharre, *America can win the AI race*, “Foreign Affairs”, 4/3/2023.

Ancora una volta nella sua storia, l'America sta dunque riuscendo a mobilitare la sua base industriale per tutelare i suoi interessi di sicurezza nazionale. E tuttavia, l'avanzata di un apparato militare-industriale 2.0 non è eticamente e politicamente neutra. L'IA è infatti una tecnologia che impatta sulla vita quotidiana dei cittadini³⁰ e, evidentemente, la mancata regolazione e l'assenza di deliberazione pubblica su queste tematiche – di fatto gestite informalmente tra aziende e burocrazie – riattivano il problema weberiano del rapporto tra *Demokratisierung* e *Bürokratisierung*, cui ora volgiamo lo sguardo.

3. Riflessività sistemica, riflessività democratica e sicurezza nazionale: un triangolo impossibile?

Max Weber, nel delineare l'idealtipo del potere burocratico, afferma che “l'amministrazione burocratica è tendenzialmente una amministrazione con esclusione dalla pubblicità [*Öffentlichkeit*]. *La burocrazia protegge dalla critica, fino a che può, quello che fa*”³¹. Al fine di mantenere il “potere reale”³², gli apparati dello Stato tendono a separarsi dall'opinione pubblica, perché “non solo ogni tipo di giustizia popolare [...] ma anche ogni specie di influenza intensiva nell'amministrazione da parte della cosiddetta ‘opinione pubblica’ [...] è di ostacolo al corso razionale [...] dell'amministrazione”³³. La capacità della burocrazia di proteggere lo Stato³⁴ e di reagire rapidamente agli eventi³⁵ non può essere irretita dai “sentimenti irrazionali”³⁶ che si generano nell'opinione pubblica: le richieste di giustizia, giustificazione o deliberazione pubblica che emergono dalla dialettica sociale interferiscono, secondo Weber, con la *Zweckrationalität* propria della *Bürokratisierung*, considerata “lo specifico mezzo per trasformare un ‘agire di comunità’ [*Gemeinschaft*] in un agire ‘sociale’ razionalmente ordinato”³⁷.

Nonostante la burocrazia sia un prodotto del moderno Stato democratico³⁸, essa tende a rendersi autonoma dalle pratiche democratiche,

³⁰ La letteratura su questo è sterminata. Si veda A. Hepp, *Deep Mediatization*, Routledge, New York 2020 e S. Lindgren (a cura di), *Handbook of critical studies of Artificial Intelligence*, Elgar, Celtigham 2023.

³¹ M. Weber, *Sociologia del potere*, PGreco, Milano 2014, p. 80. Corsivo mio.

³² Id., *Parlamento e governo*, Laterza, Roma-Bari, p. 23.

³³ Id., *Sociologia del potere*, cit., p. 64.

³⁴ *Ivi*, p. 54.

³⁵ *Ivi*, p. 57.

³⁶ *Ivi*, p. 64.

³⁷ *Ivi*, p. 74.

³⁸ Cfr., Id., *Parlamento e governo*, cit., p. 23.

sviluppando agende, liturgie e gerarchie proprie che – basate sulla *durata, l'indispensabilità e la competenza* – la rendono il vero potere della nostra epoca:

Di fronte alla crescente indispensabilità della burocrazia *statale* [...] e alla crescente posizione di potenza che ne deriva, come può essere offerta qualche garanzia che esistano potenze che tengano a freno l'enorme predominio di questo ceto d'importanza sempre crescente e lo controllino efficacemente? Come sarà possibile una democrazia, anche soltanto in questo senso limitato?³⁹

Commentando questi passaggi weberiani, Habermas coglie il problema etico-politico fondamentale: le burocrazie “acquistano autonomia mediante un *delimitazione neutralizzante rispetto alle strutture simboliche del mondo di vita*”⁴⁰. Le decisioni prese dagli apparati dello Stato – sia che esse rispondano a criteri di *Zweckrationalität* sia che esse vengano giustificate in termini di sicurezza nazionale – mettono tra parentesi “lo sfondo normativo dei nessi di azione informalmente abitudinari e regolati eticamente”⁴¹, tirandosi di fatto fuori dallo spazio dell'argomentazione, del discorso e della giustificazione generale e reciproca. La *legittimità* della burocrazia non si basa su *pratiche* democratiche o deliberative, ma sulla sua capacità di mettere in campo un “agire sociale razionalmente ordinato”, ovvero in grado di rispondere prontamente e razionalmente alle sfide che le vengono presentate. La conseguenza, di certo non neutra eticamente, è che “i risultati in base ai quali il sistema giustifica sé stesso potrebbero non essere interpretati in termini politici”⁴². Il sistema è dunque in grado di riflettere su sé stesso per migliorare i suoi risultati, ma siffatta riflessività sistemica avviene sullo sfondo della rimozione delle richieste normative che emergono dalla *Sittlichkeit*.

Se applichiamo queste intuizioni alla congiuntura che abbiamo precedentemente descritto, si vede chiaramente come nel nuovo capitalismo politico americano – categoria, non a caso, weberiana⁴³ – esistano delle

³⁹ *Ivi*, p. 37.

⁴⁰ J. Habermas, *Teoria dell'agire comunicativo*, Il Mulino, Bologna 2019, 2 Voll., Vol. II, p. 431.

⁴¹ *Ivi*, p. 433.

⁴² *Id*, *Toward a rational society*, Beacon Press, Boston, p. 112.

⁴³ Cfr., M. Weber, *Economia e Società*, Edizioni di Comunità, Milano 1961, p. 164. Per una storia del concetto nell'opera di Weber, cfr. D. Trubek, *Max Weber on Law and the Rise of Capitalism*, in A. Javier Trevino (a cura di), *The sociology of law*, Routledge, New York, 2017, pp. 220-232; S. Andreski, *Max Weber on Capitalism, Bureaucracy and religion*, Routledge, New York 2013; G.H. Mueller, *Socialism and capitalism in the work of Max Weber*, “British Journal of Sociology”, 1 (1982), pp. 151-171. Si veda anche C.J. Milhaupt, K. Pistor, *Law & Capitalism*, University of Chicago Press, Chicago 2019.

efficaci forme di riflessività sistemica – che effettivamente stanno rallentando l’ascesa cinese –, che sono però totalmente slegate da pratiche di riflessività democratica. Nello sviluppo dell’IA e nella sua *governance* ciò emerge con particolare chiarezza: l’innovazione non può essere rallentata dai “sentimenti irrazionali” dell’opinione pubblica. Chiedere *policies* più stringenti, normative che contrastino i monopoli o l’adozione di standard nell’IA è una pratica che viene immediatamente rigettata dalle techno-burocrazie americane, giacché l’innovazione viene interpretata come un gioco a somma-zero: ogni spazio lasciato libero verrà occupato dalla Cina, la quale può contare su un governo autoritario che sorveglia i suoi cittadini sin dalla nascita e che può, attraverso sussidi e altri strumenti, incanalare rapidamente la produzione e l’innovazione verso i suoi obiettivi strategici.

Non è dunque casuale che, anche quando figure teoricamente favorevoli a pratiche deliberative raggiungono posizioni di potere – come nel caso di Lina Khan, accademica, autrice di un importante studio critico sul monopolio di Amazon⁴⁴, favorevole alla regolazione dell’IA e nominata da Biden presidentessa della Federal Trade Commission (FTC) –, esse raramente riescano a far valere la loro autorità *politica* contro il “potere reale” delle burocrazie federali⁴⁵, le quali riescono a bloccare le pratiche di riflessività democratica e di regolamentazione (sul modello della normativa europea GDPR⁴⁶) ricorrendo alla nozione di sicurezza nazionale e allo “spauracchio” cinese⁴⁷.

Pensare di risolvere in questa sede questi dilemmi è estremamente complesso. E tuttavia, data l’importanza della posta in gioco, la filo-

⁴⁴ L.M. Khan, *Amazon’s Antitrust Paradox*, “The Yale Law Journal”, 3 (2017), pp. 710-805.

⁴⁵ Il caso di Khan è emblematico: il 3 maggio del 2023, pubblica un editoriale sul *New York Times* in cui sostiene la necessità di regolare l’IA. Tuttavia, solo una settimana prima aveva firmato uno *statement* interagenzia della FTC in cui, nel delineare un approccio comprensivo degli apparati americani all’IA, sosteneva che non c’era urgenza di regolarla, giacché “le tecnologie legate alle IA sono già coperte dalle leggi vigenti. Non stanno emergendo in un vuoto. Ogni agenzia federale è già dotata degli strumenti per combattere i rischi posti dall’IA”. Cfr., Id., *We must regulate AI*, “The New York Times”, 2/5/2023 e Id., *Statement of Chair Lina M. Khan Regarding the Joint Interagency Statement on AI*, “Federal Trade Commission”, 25/4/2023.

⁴⁶ Non possiamo qui approfondire il modello europeo di *governance* dell’IA. Brevemente, si può affermare che la capacità regolativa dell’Unione è resa più semplice dalla mancanza di campioni nazionali da tutelare, essendo l’infrastruttura digitale dei paesi europei di fatto costituita dalle piattaforme americane. L’Europa non ha aziende strategiche da proteggere e, dunque, non deve affrontare il dilemma che invece devono affrontare gli USA tra tutela dei cittadini e difesa della sicurezza nazionale. Su questo, si veda A. Bradford, *op. cit.*, pp. 136-145.

⁴⁷ Si veda la recentissima e dettagliatissima analisi della dialettica tra Silicon Valley e agenzie di sicurezza svolta in R.M. Shah, C. Kirchoff, *Unit X. How the Pentagon and Silicon Valley are transforming the future of war*, Scribner, New York 2024.

sofia morale non può astenersi dal porre in questione l'attuale stato di cose, senza ignorare la complessa fase storica che stiamo attraversando. Provvisoriamente, si possono immaginare due strade che l'etica può imboccare per affrontare questa congiuntura. La prima consiste nel concepirsi come pratica di "difesa critica degli spazi di autonomia pubblica"⁴⁸. Ciò non significherebbe opporsi astrattamente ai legittimi interessi di sicurezza nazionale dello Stato. Piuttosto, tale pratica avrebbe l'obiettivo di allargare lo spazio etico-normativo delle giustificazioni "a ogni livello"⁴⁹ (legislativo, burocratico ed economico), così da rendere i cittadini consapevoli degli interessi di sicurezza nazionale e le burocrazie sensibili alle altrettanto legittime preoccupazioni degli utenti per quanto riguarda questioni come la privacy e la regolamentazione dell'IA⁵⁰.

Ciò, ovviamente, potrebbe non bastare, perché presuppone una volontà politica che – al momento – pare mancare. Dunque, la seconda strada che la filosofia morale può imboccare per affrontare questi problemi etico-politici può essere quella di pensare forme di agire sociale comunicativo e solidale che sfuggano⁵¹ – in virtù della loro natura assembleare e *bottom-up* – alle pretese di "chiusura sistemica" (Habermas) che il connubio Stato-mercato-burocrazia pare generare. Anche in questo caso, l'obiettivo non sarebbe quello di negare astrattamente la congiuntura nella quale ci troviamo, ma quello di forzare l'adozione di pratiche deliberative attraverso l'innesto nella sfera pubblica di concetti eticamente e normativamente pregnanti (responsabilità, giustizia, solidarietà, rispetto) in maniera strutturale e non contingente/sensazionalistica (come avvenuto, per esempio, dopo il caso Snowden o con Cambridge Analytica).

In generale, l'obiettivo dovrebbe essere quello di mettere in campo delle pratiche deliberative che, lungi dallo schiacciarsi sul versante etico o su quello securitario, possano permettere una comprensione reciproca tra dinamiche proprie del mondo di vita e proprie del sistema, così favorendo pratiche discorsive che possano culminare con dei *trade-off* tra le diverse parti in causa, attraverso un ruolo più attivo dell'*Öffentlichkeit* nelle pratiche di *agenda setting* e definizione delle priorità. Ciò presupporrebbe, almeno a livello teorico, un tentativo di ricostruzione della forza critica del concetto di opinione pubblica, che oggi pare essere stretto tra la Scilla della rinnovata *Bürokratisierung* e la Cariddi della

⁴⁸ A. Volpe, *Dall'azione alla regolamentazione: autonomia pubblica, automazione e intelligenza artificiale*, "The Future of Science and Ethics", 1(2023), p. 68.

⁴⁹ *Ivi*, p. 67.

⁵⁰ Per una panoramica di queste preoccupazioni, cfr. M.D. Dubber, F. Pasquale, S. Das, *The Oxford Handbook of Ethics of AI*, Oxford University Press, Oxford 2020.

⁵¹ Cfr. M. Russo, *Beyond the dark side of the web: for an ethical model of digital solidarity*, "Critical Horizons", 1(2024), pp. 1-13.

frammentazione portata dai *media digitali* in termini di polarizzazione ed *echo chambers*⁵².

Peraltro, anche i report strategici americani sottolineano l'esigenza di creare un clima di fiducia nei confronti delle nuove tecnologie⁵³. In questo senso, un'etica critica dell'IA – che introduca nella sfera pubblica nozioni eticamente rilevanti e che si sforzi di difendere l'autonomia pubblica – potrebbe contribuire proprio a questo obiettivo. La filosofia morale è insomma in grado di mostrare come gli obiettivi di sicurezza nazionale possano essere raggiunti anche – se non, addirittura, in modo *migliore* – integrando i cittadini in un processo deliberativo effettivamente aperto.

4. Conclusione: per un'etica strategica

Le considerazioni sin qui svolte ci permettono di tematizzare un nuovo campo di studi che la teoria critica della società dovrebbe tenere in considerazione, soprattutto quando si rivolge allo studio delle innovazioni tecnologiche. Infatti, per quanto il cosiddetto capitalismo della sorveglianza o l'IA siano in grado, di per sé, di generare effetti di potere⁵⁴, le difficoltà nel pervenire a una *governance* delle nuove tecnologie non dipendono esclusivamente dagli interessi economici delle aziende, ma da una trasformazione strutturale dei rapporti tra Stato e mercato, centrata sulla nozione di sicurezza nazionale. In questo senso, una teoria critica dell'IA – che normativamente si ponga l'obiettivo di promuovere una *governance* emancipativa di essa, ovvero in grado di slegarla dalle dinamiche di potere a cui è soggetta e che perpetua – non può limitarsi a denunciare e mitigare le singole patologie che essa genera.

Al contrario, l'attuale congiuntura impone un'analisi di ampio respiro su quello che, provvisoriamente, può essere definito “capitalismo politico digitale”, all'interno del quale i rapporti tra Stato e mercato si sono drammaticamente complicati. Al punto che oggi, come si è visto, a essere contrari alle regolamentazioni non sono solo (ovviamente) gli attori privati, ma anche – se non soprattutto – quelli statuali. Un'analisi approfondita dell'attuale contesto geotecnologico, geoeconomico e geopolitico è dunque fondamentale per comprendere la complessità della grande trasformazione che stiamo vivendo, permettendoci così di sviluppare una teoria critica dell'esistente ben salda nella realtà. L'attuale congiuntura genera

⁵² Cfr., J. Habermas, *Storia e critica dell'opinione pubblica*, Laterza, Roma-Bari 2019 e Id., *Nuovo mutamento della sfera pubblica e politica deliberativa*, Cortina, Milano 2023.

⁵³ Questione presente sia nei report della NSCAI che in quelli del CSG.

⁵⁴ Si vedano, rispettivamente, S. Zuboff, *Il capitalismo della sorveglianza*, LUISS University Press, Roma 2019 e S. Lindgren, *Critical Theory of AI*, Wiley, London 2023.

infatti delle patologie sociali che vanno affrontate eticamente, nel tentativo – come si è cercato di accennare – di contestare ogni forma di chiusura sistemica, introducendo, a ogni livello, quelle strutture etico-normative che la nozione di sicurezza nazionale tende a mettere tra parentesi.

La scomparsa della dimensione discorsiva e deliberativa, peraltro, viene utilizzata da molti autori – non senza ragioni – per sottolineare le vicinanze, pur con le dovute differenze, tra il modello americano e quello cinese per quanto riguarda la *governance* del digitale e dell'IA: entrambi i modelli, infatti, paiono prediligere la sicurezza alla regolamentazione e alla deliberazione⁵⁵. Sotto questo riguardo, la filosofia morale può senza dubbio svolgere un ruolo fondamentale, se non addirittura strategico. Essa può – e forse deve – assumersi il compito di favorire l'autoriflessione delle potenze occidentali, le quali – impegnate nel contenimento di paesi autocratici come Cina, Russia e Iran – corrono il rischio di tralasciare quelle questioni etiche che sono inevitabilmente legate alla tradizione democratica che si vuole difendere. In questo contesto, l'etica – nel suo promuovere momenti critici di autoriflessione – avrebbe dunque il compito decisivo di riportare alla mente dei decisori occidentali le parole con cui, nel 1946, George F. Kennan ammoniva la leadership americana: “trasformarci nel problema che vogliamo risolvere è il pericolo più grande”⁵⁶.

⁵⁵ Si vedano, fra i tanti, B. Milanovic, *Capitalismo contro Capitalismo*, Laterza, Roma-Bari 2020; P.N. Wong, *Techno-Geopolitics. US-China tech war and the practice of digital statecraft*, Routledge, New York 2021; A. Bradford, *op. cit.*, pp. 57-63.

⁵⁶ G.F. Kennan, *Il Lungo Telegramma*, in G. De Ruvo (a cura di), *op.cit.*, p. 97.

*Eran Fisher**

AI as epistemic media and the future of subjectivity

Abstract

For scholars of the social sciences and humanities, the most pressing question concerning AI is its ramifications for human beings: what would it mean to be human in a world increasingly populated by AI? This, I suggest, requires us to divert from the ontological question of “what is AI?” and seek a more human-centric mode of inquiry. In this essay I would like to lay out one such possible route for inquiry. I begin by identifying the shortcomings of the common historiography of AI, which narrates a progressive development of an autonomous thinking machine from its roots in mythology up to innovations in digital technology. Instead of such history, which harks back hardly a century and culminates in the development of deep learning, I suggest thinking of AI in the context of a much longer, non-digital history of epistemic media. Epistemic media extend the cognitive abilities of human users. They facilitate the creation of new forms of knowledge about the world and about the self. For centuries now, epistemic media have offered devices for creating subjectivity by allowing reflexivity, which grew out of a dialogue between the machines and its human users. But they were not autonomous thinking machines. By thinking about AI as part of that family of media devices we can start to outline a few key axes along which questions concerning the ramifications of AI to what it means to be human can be posed.

Keywords

AI; Subjectivity; Media; Technology Discourse

1. Historicizing AI: Conflation and conflict

A certain way of historicizing AI runs the risk of mystifying it. It narrates a linear story of technological progress, where mathematical

* The Open University of Israel



breakthroughs are rendered into engineering marvels, and where smart machines become increasingly more complex, layered, scalable, and accessible. Such historiography suffers from what might be called “anachronistic teleology”. It begins with the end – fully automated thinking machines – and moves backwards in time to excavate its origins and development from this contemporary perspective. This history tends to solidify the mystique of technology by aggrandizing the people involved in its creation and ultimately tells a story of human triumph (or tragic sublime).

We can get a glimpse of the popular discourse on technology in general and AI in particular as it is registered in Wikipedia. Wikipedia entries serve here not as indications of what AI actually is, but of how they are talked about, the technology discourse which surrounds them. By technology discourse I refer to a prevailing body of knowledge which purports to make sense of society by analyzing it through the axis of technology. Critical approaches see technology discourse as an object to be analyzed in and of itself, not for its truth-value (or its adherence to the reality of technology), but for the ideological coordinates that underlie it (Mosco 2004, Fisher 2010, Nye 1994). The entry for “The history of Artificial Intelligence” underscores two bodies of knowledge that foreshadow AI: myths in antiquity about intelligent (non-human) beings and attempts by modern philosophers to describe human thinking “as the mechanical manipulation of symbols” (Wikipedia 2024a). These two bodies of knowledge, “culminated in the invention of the programmable digital computer in the 1940s”. This device, together with the ideas behind it “inspired a handful of scientists to begin seriously discussing the possibility of building an electronic brain” (Ibid.). In essence, then, this is a history of translating a fantasy and an intellectual quest into a reality by way of computers. As a research field, AI was founded in the 1950 in the U.S., went through a few decades of ups and downs, as well as a long period of disillusionment from the possibility of achieving AI, a period known as “AI Winter”. A breakthrough happened in the early 2000s with the successful application of machine learning, which relied on powerful computer hardware and the collection of big data via the internet. Eventually, in the 2010s, deep learning eclipsed all other methods and gave rise to generative AI applications.

Such historiography of AI underlies a key narrative in the popular discourse, one which situates AI *vis-à-vis* human beings. The Wikipedia entry for “Artificial Intelligence” further elucidates that, defining AI as “a field of research in computer science that develops and studies methods and software that enable machines to perceive their environment and use learning and intelligence to take actions that maximize their

chances of achieving defined goals” (Wikipedia, 2024b). A succinct representation of that discourse comes out of the key image accompanying the series of articles on AI in Wikipedia (figure 1). The image can be read as either a human brain occupied by a machine, or a machine posing as a human brain.



Figure 1: Human brain as machine/machine as human brain.
Image from Wikipedia.

Such exemplars of the discourse on AI attest to a prevalent narrative featuring a dual dynamic taking place between humans and machines: *conflation* and *conflict*. According to this, humans and AI are not only mutually exclusive, but their relations are inherently tenuous. This duality is epitomized by the most important qualifier for AI: The Turing Test. In the test, an artifice is considered intelligent to the extent that its interlocutor is not able to distinguish it from a human interlocutor. On the one hand, then, in order to be considered intelligent, the artifice needs to be perceived *as if* it were human, to be conflated with humanness to the point of indistinguishability.

On the other hand, the Turing Test is haunted by a specter of *contest* and *conflict* between humans and machines. The intelligent artifice is assumed not merely to pass as human but to eventually surpass it. It therefore threatens to make humans redundant. First, in the very practical sense, by getting rid of them in all sorts of scenarios and replace them with a computer. But more profoundly, AI threatens to dethrone humans from their superior position in the universe by bringing down the last fortress of their unique humanness: their intelligence. This would not be the first time that humans are decentered from their own cosmology. It happened once with the Copernican revolution which suggested that our earthly habitat is not the center of the universe. It happened again with Darwin, who suggested that humans are not the crown of the natural world (and possibly not the pinnacle of a divine plan), and even more radically, that they are merely a recent development in the history of nature. The substitution of a humanist cosmology for a religious cosmology shifted the gravity of agency towards human will; now it was humans who

were deemed to make their own world. But this omnipotent fantasy also received two blows which further decentered “humans” from human history: both Marx and Freud pointed to extremely robust structural constraints of human history and human instincts (respectively), which govern our humanness. In the prevalent discourse, AI is seen as posing a new threat to dethrone us a little more. Only this time, it is the result not of a discovery of natural laws, but instead the making of our own devices: we invented and developed an artifice which threatens to surpass us and control us.

AI exacerbates the contention between humanists, who insist on uniquely human qualities, which can never be fully automated (creativity, for example), and post-humanists, who cannot conceive of anything humans can do, that machines would not eventually be able to do. By post-humanists, I refer to both optimists, who imagine a bright future of smart machines enhancing human life, and pessimists, who foresee humans doomed by AI, either culturally and politically, or literally. Both the humanist and post-humanist versions are, in fact, conflictual: either humans will be able to control AI and make it subservient to their wants or be controlled by it and defeated.

Public discourse on AI, then, posits humans and machines as two distinct, discrete, contrasted entities. As AI becomes humanity’s reflection it might also become its annihilator; it is humanity’s rogue doppelgänger. Against the backdrop of this narrative, embedded in the ‘normal’ historiography of computers (to use Kuhn’s term), I want to intimate an alternative historiography by thinking of AI as yet another iteration of knowledge devices, or what I call ‘epistemic media’. This approach might open two new avenues for thinking about AI and deepen our understanding of its ramifications for humans. First, we can imagine a more dialectical history, one where humans and machines are not discrete entities but rather a mutually constitutive assemblage of knowledge production. And second, a history where the protagonists are not “humans” as such, but instead, “human subjectivity” (or self-reflection), a particular *facet* of our humanness. Such a history narrates the rise and expansion of subjectivity with the help of epistemic media during modernity and opens the question concerning the fate of subjectivity in contemporary AI-saturated environment.

2. AI as epistemic media: thinking with machines

Rather than excavating the philosophical, mathematical, and technical roots of AI, assuming a teleology of its current applications, I suggest we think of AI as one more event – albeit with seismic effects – of another

class of phenomena altogether: epistemic media. By epistemic media I mean a space, the materiality (both substance and form) of which help render data and information into a new kind of knowledge. Rather than thinking of media merely as a container of information, or a registrar of already-existing knowledge, the notion of epistemic media suggests thinking of the design and practices, involved in their use, as facilitating the construction of new knowledge. In this formulation, old, non-electronic media forms may be reimagined as calculators, or computers. Epistemic media are devices that facilitate thinking and allow for new ways of knowing to emerge. Historicizing AI as epistemic media, then, should encourage us to look for antecedents to AI, not in the technical sense but instead in the functional sense.

There is a large body of research into the most mundane and important epistemic media.² These “traditional” (i.e., non-computational) epistemic media were instrumental in bringing about modern subjectivity. By creating a space for self-reflection, where the self could encounter itself also as an object and be subjected to critique, they helped expand our realm of freedom. These media were not seen as distinct from our humanness, certainly not in conflict with it, they were perceived as facilitating a mode of knowledge co-construction. They were constructed by human beings and operated by them, but also had an agential role in affording self-reflection. An epitome of these co-constructivist relations between human subjectivity and media is the practice of diary-keeping. With modernity, self-reflection (i.e., using reason to inquire about one’s own self) came to be an integral component of the self, and was interlaced with media. This “inward turn” (Taylor 1989, 177) and the insistence on opening up a space for self-reflection was facilitated by the availability of a material space on which such reflections could be arranged. Thus, for example, the production of cheap paper in the 16th century, and the expansion of writing “had an enormous impact on the growth of the self-consciousness” (Heehs 2013, 41). Letter and diary-writing thrived in 18th century Europe. Writing was not merely an “imprint of the soul” but also a way to present the self to the self, to make it public; “the diary became

² Some of the epistemic devices that have received attention are the page (Mak, 2012), lists (Goody, 1977; Young, 2017), forms (Esbester, 2011), files (Hull, 2012; Vismann, 2008), documents (Gitelman, 2014; Riles, 2006), memos (Guillory, 2004), ledgers (Caruthers & Espeland, 1991), registrars, index cards (Krajewski, 2011), the filing cabinet (Robertson, 2021), the typewriter (Wershler-Henry, 2007), postcards (Östman 2004), diaries (Heehs, 2013; Vickery, 2021), letters, punch cards, books, newspaper clippings (Te Heesen, 2014), field-research boxes (Te Heesen, 2000), indexes (Duncan, 2021), table of contents, subject headings (Cevolini, 2022), chapters (Dames, 2023), paratext (Genette, 1997), footnotes (Grafton, 1999), notebooks (Yeo, 2014), and the word processor (Kirschenbaum, 2016).

a letter addressed to the sender” (Habermas 1991, 49). In sum, the diary was a radically new invention linking self to knowledge through media, allowing “experiments with ... subjectivity” (Ibid, 51).

3. Epistemic media: Old and new

But what makes a particular media form, and a particular media practice, qualify as epistemic media? Or better yet, assuming that all media are epistemic to a greater or lesser extent, what are the epistemic qualifiers for media? Let me lay out four distinct tenets of epistemic media twice: First, as they apply to traditional media (I am thinking mostly of the personal diary, for simplicity’s sake), and then, as they apply to AI.

1. Inscribing data: What is inscribed in epistemic media is data, as raw and naked as it gets. I understand the distinction between data, information, and knowledge as relative and co-dependent, rather than ontological (Rowley 2007). Pieces of data amount to information, pieces of information amount to knowledge. By definition, then, if epistemic media creates a new type and a new layer of knowledge out of that which was inscribed in it, all that which was inscribed can be thought of as data. Data can be the serial number of a transaction in an accounting ledger, a location of an appointment in a personal calendar, or a narrative about a dream laid out on the pages of a personal diary.
2. Real (or near real) time: Inscription takes place very close to the event being recorded. A meeting is registered in the calendar when it is set; a transaction is inscribed in the ledger when money is exchanged; an event is reported on in the personal diary not long after it had taken place. A personal diary is not a memoir, nor an autobiography, it is not a retrospective narrative with a beginning, middle, and end. Rather, it reports on events and thoughts as they happen.
3. Ordering: The spatial arrangement of data on the page is key for allowing its rendering into knowledge of a higher order.³ In a personal diary, the arrangement is chronological. Not only is such ordering *a priori*, that is, it is set before any event took place or has been recorded. More than that, chronological ordering is also an arbitrary narrative structure. One could possibly arrange entries in a personal diary by any number of categories, such as types of events (dreams, meetings, etc.), or alphabetically (e.g. representing the name of an

³ Incidentally, the French *Ordinateur*, which currently denotes a computer, is etymologically linked with order in the sense of arrangement and sorting.

entry's protagonist). Ordering can also be achieved with lines which divide the page. Double-entry bookkeeping, for example, achieves its epistemic purpose by vertical lines separating incoming from outgoing expenses. Forms do the same with rubrics.

4. Reviewing: Lastly, a particular media practice which renders media epistemic is *re-viewing* them, that is, looking back at the inscriptions with a fresh gaze. Reviewing means taking into account particular entries vis-à-vis other entries, or a combination of entries, or all entries combined. This allows for a new synthesis and a new gestalt – not merely a re-view but a new-view. A diarist does not merely re-read an entry in her personal diary. Commonly, she will reread a couple of entries, taking into account both “real-time” – when these discrete entries were written – and the sequence of entries as they unfold retrospectively. Reviewing would not merely remind her of past events but might also give her a new insight into patterns and irregularities in these events, and into her *self*, the self which had put these inscriptions on paper in real time. At this point data inscribed may become the building blocks for knowledge grasped.

Reviewing, then, can be understood as the zenith of the epistemic process which begins with data inscription, a process where self and media co-create. Or in the succinct words of Michel de Montaigne, one of the great innovators of writing as a device for reflection: “I have no more made my book than my book has made me: it is a book consubstantial with the author” (Montaigne 2003, 612). By its structural design and mode of use, epistemic media opens up a space for self-reflection, where one can render her subjective experiences into objective inscriptions, which are then reflected back to her. Epistemic media can be likened to a therapeutic space where utterances lose their status as statements of an external reality and are reconsidered to reveal an internal reality, that of a subject.

So how does AI operate as epistemic media? AI can be said to automate the epistemic process of media; facets of the epistemic process which were hitherto co-produced with human users are transferred to a fully autonomous process carried out by machines. For simplicity's sake, I will be thinking about “really-existing AI” in the following analysis, AI components embedded within digital platforms. To be even more concrete and precise, I am thinking of personalization machines poised to identify the wants and likes of individual users and marry them with relevant content, be it news stories or products.⁴

⁴ See, for example, Hallinan and Striphas 2014; Karakayali, Kostem, and Galip 2018; Prey 2018.

1. Inscribing data: The ‘intelligence’ of AI, whatever this term may stand for, begins with the collection of massive volumes and varieties of data. AI’s approach to data is omnivorous. Not unlike the writer of a personal diary, whatever the specific task of an AI machine might be, it cannot really tell which data will turn out to be relevant for the later stages of knowledge creation. It therefore opts to be as wide and persistent as possible in generating this data. Data is mostly collected automatically as users of digital platforms (either online or, increasingly, in the “real world”) leave a trail of data, registering their actions and behavior. In other words, most of the data to be inscribed is registered without the involvement of individuals (except giving consent to access the data they inadvertently create).
2. Real (or near real) time: Data collection on digital platforms is done in real time. It is unmediated, un-interpreted, “dumb” (compared with the intelligence it purports to bring about). At the moment of inscription, it carries no further weight than the event it registers. In that sense it is – by and large – non-representational; it records the binary movements that have occurred. If a click of a user on a webpage creates the command 1110010110, then the data point is 1110010110. The real time quality of this process is succinctly captured by the notion of “lively data” – data as an organic entity, always on the move, always changing – or that of “flow” (of data) as a key dynamic to contemporary social action on digital platforms (Lupton 2016; Cetina and Preda 2007).
3. Ordering: Each data point needs to be ordered in space and time. The most fundamental and prevalent axis of order is temporality. For any data point, digital media needs to register a meta-data of time signature. In that sense it is just like a calendar, a personal diary, or a ledger: each piece of data is coupled with another piece of meta-data, ordering the former. Time signature (in all these media) needs not be concrete: in a personal diary, where inscription in real time is linear, each word and phrase also carries with it a time signature (however crude it may be).
4. Reviewing: The zenith of the epistemic process is here done completely automatically and autonomously from humans. It is a process of pattern recognition, which is a statistical procedure at its essence. The power of AI is statistical. This makes it, probably, the “purest” epistemic media in that the content of the data it processes is irrelevant for the results. All the knowledge it produces is deduced from statistical information about the frequency of an event in a universe of events.

4. Conclusion

This short essay is quite modest in its scope and goal. It tells us little about the ontology of AI and does not allude to the possible ramifications of AI on humans. Instead, it seeks to offer an alternative route for inquiry by doing two things. First, transposing AI from its convenient location in the history and genealogy of information theory and technology to that of epistemic media. Such a view has the benefit of disarming our enchanted (or terrified) gaze of digital technology. It does not refute the novelty of AI, or even its revolutionary potential, but instead seeks to elucidate this novelty in light of another – existing, even old – analytical axis. Second, placing AI as an element in the set “epistemic media” also allows us to shift the discussion from a general inquiry about humans versus AI, and focus on a specific element of humanness: subjectivity.

I have argued elsewhere that certain applications of AI – particularly algorithmic devices that render user-generated data into knowledge about them, which in turn informs how users think about themselves – can be thought of as epistemic media (Fisher 2021, 2023). My argument was that in comparison with “traditional” epistemic media, AI is excluding the self from the production of knowledge about the self. But even if we accept the thrust of this argument, many questions are left open and need further exploration in order to understand the possible ramifications of AI on human subjectivity. A more robust exploration, I suggest, may be carried out using the four tenets of epistemic media outlined above.

References

- Carruthers, Bruce and Wendy Nelson Espeland. 1991. “Accounting for Rationality: Double-Entry Bookkeeping and the Rhetoric of Economic Rationality,” *American Journal of Sociology* 97(1): 30-67.
- Cetina, Karin Knorr and Alex Preda. 2007. The Temporalization of Financial Markets: From Network to Flow, *Theory, Culture & Society*, 24(7-8), 116-138.
- Cevolini, Alberto. 2022. Remembering and Forgetting by Means of Subject Headings: A Contribution to the History of Knowledge Organization, *Journal for the History of Knowledge* 3(1): 1-13.
- Dames, Nicholas. 2023. *The Chapter: A Segmented History from Antiquity to the Twenty-First Century*. Princeton: Princeton University Press.
- Duncan, Dennis. 2021. *Index, A History of the: A Bookish Adventure from Medieval Manuscripts to the Digital Age*. New York: W. W. Norton & Company
- Esbester, Mike. 2011. Taxing Design? Design and Readers in British Tax Forms before 1914. *Design Issues* 27(3): 84-97.

- Fisher, Eran. 2010. *Media and New Capitalism in the Digital Age: The Spirit of Networks*. New York: Palgrave.
- Fisher, Eran. 2021. Epistemic media and critical knowledge about the self: Thinking about algorithms with Habermas, *Critical Sociology*, 48(7-8): 1309-1324.
- Fisher, Eran. 2023. Epistemic media: Their history and relations to subjectivity, *Knowledge Cultures*, 11(3): 7-24.
- Genette, Gerard. 1997. *Paratexts: Thresholds of Interpretation*. Cambridge: Cambridge University Press
- Gitelman, Lisa. 2014. *Paper Knowledge : Toward a Media History of Documents*. Durham: Duke University Press.
- Goody, Jack. 1977. *The Domestication of the Savage Mind*. Cambridge: Cambridge University Press.
- Grafton, Anthony. 1999. *The footnote: A curious history*. Harvard University Press.
- Guillory, John. 2004. The memo and modernity. *Critical inquiry*, 31(1), 108-132.
- Habermas, Jürgen. 1991. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Cambridge: MIT Press.
- Heehs, Peter. 2013. *Writing the Self: Diaries, Memoirs, and the History of the Self*. London: Bloomsbury.
- Hallinan, Blake & Striphas, Ted. 2014. Recommended for you: The Netflix prize and the production of algorithmic culture. *New Media and Society*, 18(1), 117-137.
- Hull, Matthew. 2012. Documents and Bureaucracy. *Annual Review of Anthropology*, 41: 251-267.
- Karakayali, Nedim, Kostem, Burc, & Galip, Idil. 2018. Recommendation systems as technologies of the self: Algorithmic control and the formation of music taste. *Theory, Culture and Society*, 35(2), 3-24.
- Kirschenbaum, Matthew. 2016. *Track changes: A literary history of word processing*. Cambridge: Harvard University Press.
- Krajewski, Markus. 2011. *Paper machines: About cards & catalogues, 1548-1929*. MIT Press.
- Lupton, Deborah. 2016. *The quantified self*. Cambridge: Polity Press.
- Mak, Bonnie. 2012. *How the Page Matters*. Toronto: University of Toronto Press.
- Montaigne, Michel de. 2003. *Michel de Montaigne: The Complete Works*, trans. by Donald M. Frame. London: Everyman's Library.
- Mosco, Vincent. 2004. *The digital sublime: Myth, power, and cyberspace*. Cambridge, MA: MIT Press.
- Nye, David. 1994. *American technological sublime*. Cambridge, MA: MIT Press.
- Prey, Robert. 2018. Nothing personal: Algorithmic individuation on music streaming platforms. *Media, Culture and Society*. 40(7), 1086-1100.
- Östman, Jan-Ola. 2004. "The postcard as media", *Text & Talk*, 24(3): 423-442.
- Riles Annelis (Ed.). 2006. *Documents: Artifacts of Modern Knowledge*. Ann Arbor: University of Michigan Press.
- Robertson, Criag., 2021. *The filing cabinet: A vertical history of information*. Minneapolis: University of Minnesota Press.
- Rowley, Jennifer. 2007. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), 163-180.

- Taylor, Charles. 1989. *Sources of the Self: The Making of the Modern Identity*. Cambridge: Cambridge University Press.
- Te Heesen, Anke. 2014. *The newspaper clipping: A modern paper object*. Manchester: Manchester University Press
- Vickery, Amanda. 2021. A Self off the Shelf: The Rise of the Pocket Diary in Eighteenth-Century England, *Eighteenth-Century Studies* 54(3): 667-686.
- Vismann, Cornelia. 2008. *Files: Law and Media Technology*. Stanford: Stanford University Press.
- Wershler-Henry, Darren Sean. 2007. *The iron whim: A fragmented history of typewriting*. Ithaca: Cornell University Press.
- Wikipedia 2024a. "The history of Artificial Intelligence". Accessed September 1, 2024.
- Wikipedia 2024b. "Artificial Intelligence". Accessed September 1, 2024.
- Yeo, Richard. 2014. *Notebooks, English Virtuosi, and Early Modern Science*. Chicago: University of Chicago Press.
- Young, Liam. 2017. *List Cultures: Knowledge and Poetics from Mesopotamia to BuzzFeed*. Amsterdam: Amsterdam University Press

Fabio Fossa*

Autonomia, automazione, mediazione. Alcune note critiche sull'etica dei veicoli a guida autonoma

Abstract

Autonomy, automation, mediation. Some critical notes on the ethics of self-driving cars

This paper discusses how automated vehicles ambiguously mediate the ethical value of human autonomy. By offering a criticism of the hype surrounding the issue, it suggests that different forms of driving automation mediate different aspects of human autonomy differently. Its promotion and protection can only be incompletely achieved through a narrow focus on the design and deployment of automated vehicles. Therefore, a broader focus on transportation ethics is required.

Keywords

Ethics of Technology; Autonomy; Automated Vehicles; Mediation; Transportation Ethics¹

Molte tecnologie di Intelligenza Artificiale (IA) sembrano motivare grandi aspettative, sapientemente alimentate dalle aziende che le sviluppano, circa gli impatti positivi che saranno in grado di produrre sulla vita degli esseri umani e sulle condizioni del pianeta. In aggiunta a valori di carattere pratico o economico, non di rado l'*hype* che le circonda insiste anche su valori propriamente morali. In questo modo, la desiderabilità delle tecnologie di IA si trova ad essere giustificata anche facendo riferimento alla loro supposta *bontà* – alla loro capacità di migliorare condizioni moralmente insoddisfacenti e di apportare benefici eticamente rilevanti.

Una simile commistione di intenti commerciali ed etici si osserva in modo particolarmente evidente nell'ambito della mobilità in riferimen-

* Politecnico di Milano

¹ La presente ricerca è stata finanziata dall'Unione Europea-Next Generation EU a valere sulle risorse previste dal PNRR Missione 4 – Componente 2 (“CN-MOST-SPOKE 9”) – Investimento 1.4, CN00000023, CUP D43C22001180001.



to all'automazione della guida. Secondo un copione diffuso ormai sia in ambito aziendale che accademico¹, dai veicoli a guida autonoma ci si attende niente di meno che una rivoluzione morale del sistema dei trasporti. Grazie all'IA, le esigenze cagionate da valori come sicurezza, inclusività, sostenibilità e autonomia sembrano destinate ad essere soddisfatte come mai prima.

Uno sguardo più attento alla questione rivela presto che la meta auspicata è assai più lontana di quanto prospettato. Una rivoluzione che trasformi i sistemi di trasporto secondo valori morali richiede ben più – e forse, soprattutto nel breve periodo, ben altro² – che una rapida adozione di veicoli a guida autonoma sulla rete stradale. Allo stesso tempo, indagare più da vicino il rapporto tra l'automazione della guida e i valori morali che ci si aspetta essa supporti risulta interessante ed istruttivo dal punto di vista dell'etica dell'IA come disciplina filosofica. A tal scopo, nelle prossime pagine proverò a mettere a fuoco il modo in cui i veicoli a guida autonoma *mediano* valori etici.

Le riflessioni che seguono sono motivate da tre intenzioni. La prima è di insistere sul fatto che eventuali effetti benefici di sistemi di IA quali i veicoli a guida autonoma sono raramente semplici effetti concomitanti della loro progettazione e adozione. Al contrario, essi sono conseguibili solo se posti come obiettivi espliciti fin dalle prime fasi del *design* della tecnologia e della sua introduzione in società – il che spesso confligge con altri fini progettuali e aziendali. Presentando una narrazione dove i fini morali confluiscono pacificamente nel perseguimento dei fini progettuali e aziendali, l'*hype* è ingannevole e deve essere sottoposto a critica. La seconda intenzione è di sottolineare come, anche qualora ci si impegni ad allineare i sistemi di IA a valori morali, permangano comunque difficoltà dovute all'ambiguità della loro mediazione tecnologica. La terza è di suggerire che la gestione responsabile di tale ambiguità sia da affrontare in riferimento non tanto al singolo sistema di IA, quanto al più ampio contesto pratico e normativo a cui appartiene. Nel nostro caso, un'inquadratura troppo stretta sull'automazione della guida rischia di perdere di

¹ Cfr. F. Fossa, *Ethics of Driving Automation. Artificial Agency and Human Values*, Springer, Cham 2023, <https://doi.org/10.1007/978-3-031-22982-4>

² Sulla tesi, ripresa in conclusione di questo scritto, per cui l'etica dei veicoli autonomi dovrebbe essere discussa nel più ampio contesto di un'etica dei trasporti, si vedano F. Santoni de Sio, *The European Commission report on ethics of connected and automated vehicles and the future of ethics of transportation*, in "Ethics and Information Technology", n. 23, 2021, pp. 713-726, <https://doi.org/10.1007/s10676-021-09609-8>; B. Kock, Y. Lannquist, *Fulfilling the Promise of Autonomous Vehicles with a New Ethics of Transportation*, in R. Jenkins, D. Cerný, T. Hríbek (a cura di), *Autonomous Vehicle Ethics: The Trolley Problem and Beyond*, Oxford Academic, New York 2022, pp. 390-414, <https://doi.org/10.1093/oso/9780197639191.003.0023>

vista il contesto più ampio della mobilità, in cui si radicano le difficoltà morali a cui la tecnologia vuole rispondere. Le questioni etiche sollevate dalla guida autonoma assumono la loro corretta dimensione solo nel contesto di un'etica dei trasporti.

Animate da queste intenzioni, le prossime pagine esaminano più da vicino il modo ambiguo in cui il valore morale dell'autonomia umana viene mediato nell'automazione della guida. La complessa relazione tra autonomia umana e automazione tecnologica è al centro del dibattito non solo nel caso della guida autonoma, ma anche di altre tecnologie di IA³, come ad esempio i sistemi d'arma autonomi⁴ e i sistemi di raccomandazione⁵. Non potrebbe essere altrimenti. Nella cultura occidentale, l'autonomia gode di ampio riconoscimento sociopolitico come valore etico fondamentale. In estrema sintesi, essa caratterizza gli esseri umani come enti capaci di autodeterminazione, pur se esposti a condizionamenti di diversa natura e immersi in molteplici contesti relazionali. In quanto tali, meritano rispetto e protezione⁶. L'autonomia sta alla base di aspetti tanto cruciali della vita morale quali l'esercizio della responsabilità e il pieno godimento della dignità umana. Non può, quindi, che qualificarsi come un valore chiave da rispettare e perseguire anche attraverso l'innovazione tecnologica – automazione della guida inclusa.

Per le stesse ragioni, una tecnologia potenzialmente in grado di ampliare o agevolare l'esercizio dell'autonomia umana ben si presta ad essere promossa come *buona*. Rimandi all'autonomia sono comuni nell'*hype*, il quale raramente manca di evidenziare come l'automazione della guida avrà effetti rivoluzionari sul suo godimento. Seguendo il copione a cui si è accennato sopra, innanzitutto la guida autonoma permetterà di riscattare tempo ed energie prima sprecate a guidare, le quali potranno così essere impiegate in attività scelte liberamente. Ciò determinerà anche un accrescimento dell'autonomia decisionale su questioni che influenzano significativamente il benessere individuale – come, ad esempio, dove vivere, studiare, lavorare – ora ristrette in modo tangibile dalla capacità di guidare e dalla lunghezza del tragitto da percorrere ogni giorno. Inoltre,

³ Cfr. F. Fossa, V. Schiaffonati, G. Tamburrini (a cura di), *Automi e persone. Introduzione all'etica dell'intelligenza artificiale e della robotica*, Carocci, Roma 2021; A. Laitinen, O. Sahlgren, *AI Systems and Respect for Human Autonomy*, in "Frontiers in Artificial Intelligence", n. 4, 2021, pp. 1-14, <https://doi.org/10.3389/frai.2021.705164>

⁴ Cfr. G. Tamburrini, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Carocci, Roma 2020.

⁵ Cfr. T. Numerico, *Big Data e algoritmi. Prospettive critiche*, Carocci, Roma 2020; S. Tiribelli, *Identità personale e algoritmi. Una questione di filosofia morale*, Carocci, Roma 2023.

⁶ Cfr. J. Christman, *Autonomy in Moral and Political Philosophy*, in E. Zalta (a cura di), *The Stanford Encyclopedia of Philosophy (Fall 2020 Edition)*. <https://plato.stanford.edu/archives/fall2020/entries/autonomy-moral/>

i veicoli a guida autonoma offriranno una nuova opportunità di trasporto privato a persone che ora non ne possono beneficiare per via di disabilità cognitive o fisiche, età o preferenze personali, migliorando così notevolmente la loro capacità di essere in controllo della propria vita e di prendere decisioni autonome.

A giudicare da quanto appena detto, non sembrerebbe esserci alcuna ambiguità nel modo in cui l'automazione della guida media il valore dell'autonomia. La mediazione pare del tutto positiva: *essendo come sono* – così si sostiene – i veicoli a guida autonoma implicano una concomitante estensione dell'esercizio dell'autonomia. Uno sguardo critico che interroghi più nel dettaglio aspetti sia tecnici che normativi rivela invece una situazione più ambigua. Ausili e ostacoli all'esercizio dell'autonomia sono vari, diversamente intrecciati tra loro, e frutto di esplicite scelte di progettazione. Se decisioni progettuali responsabili e disposizioni politiche efficaci dipendono in gran parte da un quadro rigoroso delle opportunità e dei rischi, un affondo filosofico sulla mediazione dell'autonomia umana nell'automazione della guida risulta necessario.

Per riportare in superficie l'ambiguità sommersa dall'*hype*, un buon punto di partenza è l'approccio europeo all'etica della guida autonoma – approccio che rispecchia ampiamente la posizione centrale assunta dal valore dell'autonomia nella cultura occidentale. Nel 2020, un gruppo interdisciplinare di quattordici esperti nominati dalla Commissione Europea ha redatto il rapporto *Ethics of Connected and Automated Vehicles. Recommendations on Road Safety, Privacy, Fairness, Explainability and Responsibility*⁷. Il documento presenta principi e raccomandazioni mirate a supportare le parti interessate nello sforzo di allineare l'automazione della guida ai valori etici più rilevanti. In stretta connessione con l'approccio europeo all'IA affidabile⁸, il rapporto si apre con l'identificazione e la definizione dei fondamenti morali che ispirano l'intera proposta. Riconoscendone la rilevanza, gli autori includono l'autonomia tra gli otto principi etici fondamentali per l'automazione della guida, insieme a non maleficenza, beneficenza, dignità, responsabilità, giustizia, solidarietà e deliberazione inclusiva.

Il principio dell'autonomia, spiegano gli autori del documento, afferma che gli esseri umani devono essere concepiti come “agenti morali liberi”⁹, il cui diritto all'autodeterminazione esige rispetto. In relazione

⁷ European Commission: Directorate-General for Research and Innovation, *Ethics of connected and automated vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility*, Publications Office, 2020. <https://doi.org/10.2777/035239>

⁸ Commissione europea, Direzione generale delle Reti di comunicazione, dei contenuti e delle tecnologie, *Orientamenti etici per un'IA affidabile*, Ufficio delle pubblicazioni, 2019. <https://data.europa.eu/doi/10.2759/640340>

⁹ *Ivi*, p. 22. Tutte le traduzioni da questo documento sono mie.

alla guida autonoma, il principio dell'autonomia impone che i sistemi di guida siano progettati in modo tale da "proteggere e promuovere la capacità degli esseri umani di prendere decisioni circa i propri spostamenti e, più in generale, di determinare da sé valori e fini per soddisfare diverse concezioni di 'vita buona'"¹⁰. Di conseguenza, l'autonomia gioca un ruolo cruciale in diverse raccomandazioni, che spaziano dalla protezione della privacy alla promozione della capacità decisionale dell'utente e alla riduzione dell'opacità algoritmica.

Dal punto di vista etico, l'enfasi sulla protezione e la promozione dell'autonomia umana nel contesto dell'automazione della guida appare appropriata. Negare autonomia di decisione e di azione attraverso sistemi di guida autonoma rischia di generare situazioni in cui aspetti cruciali della vita dei singoli vengono determinati da altri attori (informatici, ingegneri, produttori, e così via). Questo stato di cose è chiaramente incompatibile con il diritto individuale all'autodeterminazione in questioni personali e deve essere evitato nella progettazione e nell'introduzione dei veicoli autonomi nei sistemi di trasporto.

Oltre ad esigere rispetto intrinseco, la protezione dell'autonomia umana nell'ambito dell'automazione della guida è anche importante per sostenere altri valori morali rilevanti. Ad esempio, considerare gli esseri umani come agenti morali liberi significa, allo stesso tempo, ritenerli responsabili nella misura in cui possono esercitare tale libertà. Questo è un presupposto necessario per stabilire come distribuire la responsabilità con giustizia quando si verificano conseguenze dannose derivanti dall'uso dei veicoli autonomi¹¹.

Se l'importanza del valore dell'autonomia umana per l'automazione della guida non è ambigua, non vale lo stesso per la sua applicazione. Qui, al contrario, sorgono difficoltà non indifferenti. L'automazione dell'atto umano di guidare, d'altronde, implica la sua delega parziale o totale a sistemi di IA, che ne mediano l'esecuzione. Una qualche limitazione dell'autonomia dell'individuo, che da guidatore si tramuta in supervisore o semplice passeggero, non può che verificarsi. Come in ogni altro caso di mediazione tecnologica¹², anche quando l'atto della guida viene mediato da sistemi di IA l'esperienza umana ne esce trasformata in modo ambiguo. Alcuni suoi aspetti ne sono amplificati o agevolati, altri

¹⁰ *Ibidem*.

¹¹ S. Nyholm, *The Ethics of Crashes in Self-Driving Cars: a Roadmap II*, in "Philosophy Compass", a. 13, n. 7, 2018, pp. 1-10, <https://doi.org/10.1111/phc3.12506>

¹² Sul tema della mediazione tecnologica, messo a fuoco nella tradizione della postfenomenologia, si vedano D. Ihde, *Technology and the Life World. From Garden to Earth*, Indiana University Press, Bloomington-Indianapolis 1990; P.-P. Verbeek, *What Things Do. Philosophical Reflections on Technology, Agency, and Design*, Penn State University Press, University Park 2005.

ne sono ridotti o ostacolati. Nella misura in cui l'autonomia costituisce un aspetto di questa esperienza, il suo esercizio ne uscirà ugualmente trasformato. Bisogna perciò valutare quali aspetti dell'autonomia umana siano agevolati o ostacolati, e quanto, quando la guida viene mediata da sistemi di IA.

La definizione di autonomia avanzata nel rapporto europeo è di aiuto per comprendere più precisamente ciò che è in gioco in questo contesto. Essa combina due aspetti dell'esercizio dell'autonomia al volante. Il primo aspetto riguarda l'esercizio del controllo individuale sulle decisioni relative ai propri spostamenti. Nel caso della guida, ciò riguarda i modi in cui il veicolo si sposta. Per fare riferimento al celebre modello di John Michon¹³, sono qui comprese le decisioni di guida di carattere *operativo* e *tattico* – che, cioè, hanno a che vedere con il controllo del veicolo e la navigazione sulla rete stradale. Esempi sono quando e come accelerare o frenare; a che velocità andare; quanto spazio lasciare tra il veicolo e utenti vulnerabili della strada come pedoni, ciclisti o motociclisti; se dare o meno la precedenza o fermarsi presso un attraversamento pedonale; se fare uno strappo al codice della strada per un qualche fine superiore, e così via. In questo senso, rispettare l'autonomia degli utenti significherebbe permettere loro di esercitare un certo grado di controllo sulle operazioni che incidono sul modo in cui il veicolo a guida autonoma si sposta sulla strada.

Il secondo aspetto, invece, ha una portata più ampia. Riguarda i sistemi di trasporto come mezzi fondamentali per il perseguimento della vita buona, pur nella pluralità delle concezioni che sono moralmente ammissibili a riguardo. Tornando al modello di Michon, la definizione sembra ora riferirsi alle decisioni di guida di tipo *strategico*, che hanno a che fare non tanto con il controllo del veicolo, ma con il fine per cui lo si adopera – le quali appartengono al più generale contesto pratico delle decisioni su come vivere bene. Decisioni di questo genere riguardano ad esempio la meta da raggiungere, i motivi per cui la si vuole raggiungere, il percorso che si preferisce fare per raggiungerla, se andarci da soli o con altri, e così via. In questo senso, allineare l'automazione della guida al principio di autonomia significherebbe concepirla come strumento per supportare la ricerca autonoma della vita buona.

Quando guidiamo, quindi, esercitiamo la nostra autonomia a più livelli: nel modo in cui operiamo il veicolo, nelle decisioni di guida che prendiamo mentre siamo al volante, nelle scelte se andare dove, quando, come, perché, e con chi. L'atto della guida comporta un esercizio sfac-

¹³ J.A. Michon, *A Critical View of Driver Behavior Models: What Do We Know, What Should We Do?*, in L. Evans, R.C. Schwing (a cura di), *Human Behavior and Traffic Safety*, Springer, Boston 1985, pp. 485-524, https://doi.org/10.1007/978-1-4613-2173-6_19

ceffato dell'autonomia. L'automazione della guida media questi diversi esercizi in modo diverso. Non solo: anche l'automazione della guida non è una, ma è molteplice. Diversi livelli di automazione¹⁴ agevolano e ostacolano in modo diverso i diversi livelli di autonomia che vengono esercitati nell'atto della guida. Il panorama che si offre alla vista è ambiguo. Si comincia a intuire la complessità del dovere, così chiaro dal punto di vista normativo, di implementare il principio dell'autonomia nei veicoli autonomi.

Vediamo la questione un po' più da vicino. Iniziamo considerando come l'automazione della guida media l'esercizio dell'autonomia umana sugli spostamenti del veicolo, cioè sulle decisioni operative e tattiche. Dal momento che l'esecuzione dei compiti operativi e tattici è delegata al sistema di guida, l'esercizio di questo genere di autonomia non può che essere variamente ostacolato o ridotto. Per quanto alcune decisioni operative e tattiche non siano di interesse morale, altre chiamano in causa valori etici. Ostacoli e riduzioni di questo tipo sorgono ad ogni livello di automazione della guida. Anche un semplice sistema di controllo della velocità che non permetta di guidare oltre al limite imposto per legge e non possa essere disattivato dal guidatore, nemmeno in caso di emergenza, potrebbe essere considerato un ostacolo rispetto all'esercizio dell'autonomia¹⁵. All'estremo opposto, si ipotizzi l'esistenza di veicoli completamente autonomi le cui logiche di controllo siano in grado, nel caso di collisioni inevitabili, di distribuire i danni secondo valori etici predefiniti. Se tali valori fossero stabiliti non dagli utenti, ma da altri soggetti, la loro autonomia morale sarebbe ostacolata o ridotta¹⁶. Considerando casi meno futuristici, si pensi alla scelta di dimostrare (o non dimostrare) rispetto e cura guidando con estrema prudenza vicino a utenti della strada vulnerabili come ciclisti o pedoni; o alla scelta di adottare (o non adotta-

¹⁴ I livelli di automazione della guida (L0: no automation, L1: driving assistance, L2: partial, L3: conditional, L4: high, L5: full driving automation) sono stati definiti dalla Society of Automotive Engineers (SAE) nel documento J3016, intitolato *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Il documento, ampiamente adottato, è disponibile al seguente link: https://www.sae.org/standards/content/j3016_202104/

¹⁵ Cfr. J. Schoonmaker, *Proactive Privacy for a Driverless Age*, in "Information & Communications Technology Law", 25, n. 2, 2016, pp. 96-128, <https://doi.org/10.1080/13600834.2016.1184456>; J. Smids, *The Moral Case for Intelligent Speed Adaptation*, in "Journal of Applied Philosophy", n. 35, 2018, pp. 205-221, <https://doi.org/10.1111/japp.12168>

¹⁶ Cfr. G. Contissa, F. Lagioia, G. Sartor, *La manopola etica: i veicoli autonomi eticamente personalizzabili e il diritto*, in "Sistemi intelligenti", n. 3, 2017, pp. 601-614, <https://doi.org/10.1422/88511>; J. Millar, *Ethics Settings for Autonomous Vehicles*, in P. Lin, K. Abney, R. Jenkins (a cura di), *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. Oxford University Press, Oxford 2017, pp. 20-34, <https://doi.org/10.1093/oso/9780190652951.003.0002>

re) stili di guida cosiddetti ‘ecologici’, volti a moderare le emissioni. L’automazione di questi aspetti della guida implica un vincolo all’esercizio umano dell’autonomia. La sua problematicità diventa evidente quando le logiche di controllo del veicolo determinano stili di guida diversi da quelli che l’utente, se fosse in controllo, si sentirebbe tenuto a esercitare.

Alla luce di quanto sopra, sembra ragionevole concludere che la corsa verso l’automazione non dovrebbe ostacolare forme di controllo umano sui compiti di guida che permettano l’esercizio legittimo dell’autonomia. Come suggerisce la teoria del *Controllo Umano Significativo*¹⁷, se alcune decisioni di guida spettano agli utenti, i sistemi di IA devono permettere che siano prese in modo autonomo. Al livello della piena automazione, in cui il sistema di guida è esplicitamente progettato per non richiedere alcun tipo di supervisione e controllo da parte del passeggero, difficilmente queste esigenze possono essere soddisfatte. Una forma minima di controllo potrebbe essere realizzata soltanto indirettamente, ad esempio tramite la configurazione delle preferenze del sistema. Tuttavia, questa soluzione non soddisfa appieno le esigenze del principio di autonomia. Sembra più convincente affermare che il rispetto dell’autonomia inteso come controllo sulle decisioni di guida operative e tattiche sia ostacolato e ridotto dalla piena automazione.

Al contrario, livelli di automazione parziale o condizionale della guida, dove guidatori umani e sistemi di IA condividono il controllo sulle decisioni operative e tattiche, sembrano lasciare spazio sufficiente all’esercizio dell’autonomia degli utenti. La mediazione dell’esperienza della guida che è loro propria, si potrebbe argomentare, non riduce né ostacola necessariamente il controllo su decisioni operative e tattiche. Al contrario, nel caso in cui offra assistenza mirata contro distrazione, stanchezza, stress o limitazioni percettive, l’automazione parziale o condizionale può persino agevolare l’esercizio dell’autonomia, contrastando limitazioni a cui è esposta. Si pensi, ad esempio, a un sistema di guida che aiuti il guidatore a non superare un limite di velocità imposto dal guidatore stesso anche qualora si distraiga, o a un sistema che segnali la presenza di ciclisti o pedoni negli angoli ciechi del veicolo e, se necessario, inviti il guidatore a occupare diversamente la carreggiata o a seguire traiettorie più sicure.

E tuttavia, l’ambiguità rimane. La conclusione per cui l’autonomia personale sarebbe meglio tutelata da livelli di automazione parziale o condizionale, dove umani e sistemi di IA condividono il controllo della guida, diventa controversa se il valore è inteso in riferimento alla secon-

¹⁷ F. Santoni De Sio, J. van den Hoven, *Meaningful human control over autonomous systems: A philosophical account*, in “Frontiers of Robotics and AI”, a. 5, n. 15, 2018, pp. 1-14, <https://doi.org/10.3389/frobt.2018.00015>

da accezione che abbiamo enucleato. La piena automazione può supportare il perseguimento autonomo di diverse concezioni di vita buona meglio di quanto possa fare l'automazione parziale o condizionale. Almeno due opportunità si distinguono: l'inclusività e la riappropriazione del tempo di viaggio. Entrambe abilitano in modo significativo e trasversale il perseguimento di concezioni diverse di vita buona. Entrambe non sono realizzabili se ci si affida a modelli di automazione parziale o condizionale.

Da un lato, la piena automazione potrebbe migliorare notevolmente l'autonomia di categorie sociali attualmente escluse dalla guida manuale a causa di disabilità fisiche e cognitive – individui per i quali l'accesso indipendente e privato ai trasporti impatta in modo cruciale sul benessere personale e la qualità della vita. Poiché i compiti di guida sarebbero automatizzati, le disabilità fisiche e cognitive non costituirebbero più una barriera insormontabile all'uso autonomo dei veicoli privati. Questo punto, che non sfugge al copione dell'*hype*, non è però così pacifico come sembra¹⁸. Individui con disabilità fisiche o cognitive incontreranno necessariamente altre difficoltà nell'utilizzo autonomo anche del più avanzato sistema di guida, se questo non sarà interamente progettato partendo dalle loro peculiarità. Automatizzare la guida non basta a renderla inclusiva: servono scelte di progettazione specifiche, che si facciano carico delle esigenze specifiche proprie di queste categorie di utenti. Un veicolo autonomo progettato per il cliente qualunque è tanto escludente per chi non rientra in questa categoria quanto un veicolo a guida manuale. E come è ovvio, progettare un veicolo autonomo che sia davvero inclusivo richiede impegno e risorse certamente superiori rispetto a distribuire lo stesso modello a tutti e lodarne l'inclusività come effetto concomitante. Al di là delle storture dell'*hype*, rimane pur vero che la piena automazione della guida unita a un processo di progettazione attento ai bisogni specifici di date categorie di utenti potrebbe sì offrire loro, almeno in linea di principio, un'amplificazione estremamente significativa dell'esercizio dell'autonomia.

Dall'altro lato, e in modo affatto trasversale, la piena automazione della guida potrebbe facilitare il perseguimento autodeterminato di una vita buona permettendo agli utenti di non sprecare il tempo di viaggio impegnati forzatamente nella conduzione del veicolo. Liberati dalla schiavitù della guida, gli utenti dei veicoli autonomi potrebbero impiegare il tempo di viaggio come preferiscono. Una liberazione che può espandersi in susseguenti effetti liberatori su altri aspetti pratici che impattano la capacità

¹⁸ Cfr. G. Goggin, *Disability, Connected Cars, and Communication*, in "International Journal of Communication", n. 13, 2019, pp. 2748-2773, <https://ijoc.org/index.php/ijoc/article/view/9021>

degli individui di inseguire proprie concezioni di vita buona, facilitando decisioni autonome su questioni tanto importanti quali dove vivere, dove lavorare, dove curarsi, dove studiare – tutte decisioni che sarebbero meno vincolate dalla considerazione del tempo perso a viaggiare da un luogo all'altro.

In entrambi i casi sopra descritti, i benefici in termini di autonomia umana dipendono interamente dalla piena automazione della guida. Di fatto, gli individui esclusi dalla guida manuale sarebbero anche candidati poco adatti al controllo condiviso. Allo stesso modo, la piena automazione è necessaria affinché gli utenti dei veicoli autonomi possano dedicarsi liberamente ad altre attività più soddisfacenti. La tendenza degli utenti a distrarsi e non supervisionare adeguatamente il funzionamento dei sistemi di guida a livelli di automazione intermedi è già tristemente nota per aver causato incidenti mortali. Perché agevoli l'autonomia come libertà di perseguire la propria concezione di vita buona, l'automazione della guida deve progressivamente eliminare il bisogno dell'intervento e della supervisione umana.

L'ambiguità dei modi in cui diversi modelli di automazione della guida mediano diversi aspetti dell'autonomia umana è evidente. La piena automazione amplifica l'autonomia nel perseguimento di obiettivi e valori di vita riducendo il controllo su decisioni operative e tattiche. L'automazione parziale o condizionale agevola l'autonomia nel controllo del veicolo, ma ne ostacola l'esercizio in termini strategici. Il dovere di promuovere e proteggere il principio etico dell'autonomia personale può seguire direzioni difficili da armonizzare. L'ambiguità deriva dalla complessità della nozione di autonomia e da modelli divergenti di automazione della guida, che ne mediano in modo diverso aspetti diversi. Questa incertezza lascia aperta la domanda su come incorporare il valore dell'autonomia nell'automazione della guida.

La tensione che ne deriva pone una questione etica molto delicata. Allineare l'automazione della guida a legittime aspettative morali in termini di autonomia personale è un compito importante. Violazioni in questo senso rischiano di generare dolore, offese, sfiducia e reazioni negative da parte del pubblico. Tuttavia, la natura ambigua della mediazione rende complicato il passaggio da approvazioni astratte a raccomandazioni pratiche per la progettazione e le politiche.

Cosa fare nei confronti di questa ambiguità? Invece di pretenderne l'eliminazione *by design* o di nascerla sotto il tappeto dell'*hype*, bisogna riconoscerla ed imparare a gestirla. Se a livello di sistema di guida la situazione è irrimediabilmente ambigua, si deve anche allargare la prospettiva. Amplificazioni e riduzioni dell'autonomia mediate dall'automazione della guida, d'altronde, non esauriscono di certo la questione di come il valore dell'autonomia umana possa essere messo a rischio o protetto e

promosso nelle nostre interazioni quotidiane con tecnologie di trasporto. Al contrario, rappresentano solo un dettaglio di un quadro assai più ampio e variopinto. Calare il discorso sull'autonomia morale come valore cruciale per l'automazione della guida nel più ampio contesto, a cui appartiene, di una riflessione etica sulla mobilità in generale¹⁹ – nel contesto, cioè, di un'*etica dei trasporti* – potrebbe fornire strumenti efficaci per gestire la questione sollevata nelle righe precedenti e permetterebbe di comprendere meglio le istanze che la motivano.

¹⁹ Allo stesso tempo, ciò implica considerare anche un ulteriore livello di autonomia umana, ovvero il suo esercizio pubblico in relazione a decisioni che riguardano non solo l'automazione della guida, ma la mobilità in generale. Su ciò si veda ad es. F. Santoni de Sio, F. Fossa, *Designing Driving Automation for Human Autonomy: Self-Determination, the Good Life, and Social Deliberation*, in F. Fossa, F. Cheli (Eds.), *Connected and Automated Vehicles. Integrating Engineering and Ethics*, Springer, Cham 2023, pp. 19-36. https://doi.org/10.1007/978-3-031-39991-6_2. Ringrazio un revisore anonimo per aver sollevato questo punto.

*Pasquale Grieco**

Design for action: sviluppo tecnologico e metamorfosi delle categorie etiche

Abstract

Design for action: technological development and metamorphosis of ethical categories

This paper explores key conceptual intersections between subjectivity, technological mediation, and agency, focusing on the influence of information technologies, artificial intelligence, and digital systems on these foundational philosophical and ethical categories. The objective is to offer critical insights for rethinking these concepts in the context of the digital revolution. In light of the design turn, the study emphasizes the growing importance of embedding ethical values into artifacts and design processes, highlighting their implications for political and legislative decision-making at both national and supranational levels.

Keywords

Design Ethics; Artificial Intelligence Ethics; Technology Ethics; Digital Design; Digital Ethics

1. HAL e i pappagalli: l'etica nella società dell'informazione

Le domande paradigmatiche della *computer ethics*, di cui può essere considerata emblema quella di Daniel Dennett, “*When HAL kills, who's to blame?*” (Dennett, 1997), un tempo relegate a una dimensione fantascientifica, affascinante ma essenzialmente astratta, stanno diventando sempre più concrete nella nostra quotidianità. Immaginiamo, ad esempio, di chiedere un consiglio medico online e di ricevere una risposta utile per il nostro benessere. È rilevante sapere se tale risposta provenga da un medico in carne e ossa (o meglio, di carbonio) o da un chatbot (digitale, di silicio) basato su un *Large Language Model* (LLM) open-

* Università degli Studi di Bari Aldo Moro



source? La questione si complica ulteriormente, sia dal punto di vista etico che filosofico, nel caso in cui l'output risulti dannoso per la salute. Dove risiede la responsabilità morale se il suggerimento è fornito da un "pappagallo stocastico" (Bender et al., 2021), uno dei diffusi programmi di intelligenza artificiale che utilizzano il linguaggio naturale? Chi o cosa deve essere considerato responsabile in tali circostanze? A queste domande si aggiungono ulteriori interrogativi, riguardanti il ruolo della tecnologia digitale come medium dell'azione moralmente significativa.

Il modo in cui queste tecnologie vengono progettate e implementate ha una rilevanza etica? Le tradizionali categorie morali e filosofiche sono sufficienti per affrontare queste nuove tipologie di problemi? Rispondere a tali interrogativi significa confrontarsi direttamente con l'impatto etico delle tecnologie digitali e dell'informazione, in particolare delle *Information and Communication Technologies* (ICT), tra cui l'Intelligenza Artificiale (AI). Quest'ultima rappresenta, per diffusione, disponibilità e accessibilità, una potente e nuova "riserva di capacità di agire a portata di mano" (Floridi, 2022, p. 53). Spesso, le capacità di agire delle tecnologie dotate di algoritmi di *Machine Learning* (ML) e *Deep Learning* (DL) si manifestano attraverso caratteristiche quali *interattività*, *adattabilità* e *autonomia*¹.

Queste tre caratteristiche costituiscono il livello minimo necessario per considerare un agente capace di produrre effetti morali, indipendentemente dalla sua natura umana, artificiale o ibrida. Infatti, se un agente è interattivo, adattabile e autonomo, può partecipare al *moral game*² (Allen

¹ Le tre caratteristiche rappresentano il Livello di Astrazione (LoA) minimale per considerare gli agenti artificiali come agenti morali. Un LoA si riferisce a un insieme di proprietà rilevanti (osservabili, cioè varianti tipizzate interpretate) di un sistema che vengono selezionate per descrivere, analizzare e comprendere tale sistema in un determinato contesto. Cfr Floridi, 2020. In generale, sul concetto di Livello di Astrazione. Nello specifico, le tre caratteristiche di: Interattività (1), Autonomia (2), Adattabilità (3) significano: (1): che l'agente e il suo ambiente possono agire l'uno sull'altro. Esempi tipici sono l'immissione o l'emissione di un valore, o l'esecuzione simultanea di un'azione da parte dell'agente e del paziente, come la forza gravitazionale tra corpi; (2): che l'agente è in grado di cambiare stato senza rispondere direttamente all'interazione. risposta diretta all'interazione: può eseguire transizioni interne per cambiare il proprio stato. Quindi un agente deve avere almeno due stati. Questa proprietà conferisce all'agente un certo grado di complessità e di indipendenza dall'ambiente; (3): che le interazioni dell'agente possono modificare le regole di transizione con cui cambia stato. Questa proprietà garantisce che un agente possa essere visto, a un dato LoA, come un agente che apprende il proprio modo di operare in un modo che dipende criticamente dalla sua esperienza. Si noti che se le regole di transizione di un agente sono memorizzate come parte del suo stato interno, discernibile a questo LoA, l'adattabilità deriva dalle altre due condizioni. Cfr. Floridi, Sanders, 2004.

² Allen et al. hanno proposto la possibilità di un test per valutare la capacità morale delle macchine ispirato al test di Turing, *l'imitation game*. Il test morale proposto per le macchine consiste nel valutare se un agente artificiale possa superare un "*Moral Turing*

et al., 2000), anche se non è un essere umano, contrariamente a quanto previsto dalla tradizione etica della modernità filosofica.

L'apertura alla considerazione morale delle azioni di entità artificiali o ibride è di grande attualità, vista la crescente pervasività di tali tecnologie e il ruolo che rivestono nello spazio pubblico delle democrazie. Si pensi, ad esempio, agli effetti deleteri delle *fake news*, spesso propagate da *bot* sui social media, che influenzano l'orientamento del voto. Diventa quindi essenziale sviluppare una prospettiva etica sulle ICT che non solo incida sulla vita quotidiana degli individui, ma che contribuisca anche a costruire un quadro di riferimento, un *framework* morale, entro il quale gli organismi democratici (a livello nazionale, comunitario e globale) possano regolamentare la progettazione, l'implementazione e l'uso di tali tecnologie.

Il ruolo delle macchine e degli altri agenti non umani è cruciale per comprendere la società odierna, e ciò è particolarmente vero in un contesto in cui l'*informazione*³ non è solo il bene economico per eccellenza delle società capitalistiche avanzate, né soltanto il concetto chiave per interpretarne le dinamiche socio-politiche e culturali, ma rappresenta anche il livello ontologico minimo comune a tutte le entità che vi operano (cfr. Floridi, 2009, pp. 25-57). L'*agency* di macchine, agenti artificiali, sistemi ibridi e quella mediata dalle tecnologie assume dunque un ruolo

Test" (MTT), in cui si verifica se l'agente è in grado di comportarsi in modo moralmente indistinguibile da un essere umano, per un giudice umano in situazioni etiche significative. Allen, et. al., 2000. La questione che emerge è quella se è possibile considerare accettabile che le macchine debbano essere in grado di mentire riguardo questioni che hanno rilevanza etica, riguardante cioè la loro capacità di considerare e produrre bene e male, e quindi eventuali danni ad altri enti, compreso l'essere umano. Questa idea che nella possibilità di mentire vi sia la base per la capacità degli agenti artificiali di sviluppare una moralità di tipo umano si ritrova anche nel summenzionato saggio di Dennett su HAL. Posto che le macchine per sviluppare una loro eticità debbano accedere a una intenzionalità di ordine superiore ("high-order intentionality"), il prezzo da pagare per questa intenzionalità superiore è la duplice capacità di ingannare: di ingannare gli altri e se stessi. "Another price we pay for higher-order intentionality is the opportunity for duplicity, which comes in two flavors: self-deception and other-deception". Dennett, 1997.

³ L'informazione è un concetto che può essere inteso in diversi modi. Per *informazione* si può intendere l'informazione di Shannon, di natura quantitativa, che viene considerata indipendentemente dal contenuto e dalla dimensione semantica che essa può assumere. L'informazione di Shannon è definita in relazione alla misura dell'entropia e dell'incertezza di un messaggio e può essere espressa tramite il logaritmo medio dell'improbabilità del messaggio stesso, formalizzata come: $H = -\sum p_j \log_2 p_j$ (H = quantità di informazione, l'entropia; Σ =somma di tutti i messaggi j; p_j = probabilità del messaggio j). C.E. Shannon, 1948, L'informazione può essere considerata anche nella sua dimensione semantica, intesa in senso "forte" o meno. Per il dibattito sulla teoria dell'informazione semantica in senso forte si può fare riferimento a Allo, 2010. Un'introduzione efficace, accessibile e non specialistica al concetto di informazione, che ne ricostruisce anche la storia, si trova in Gleick, 2012. Per una panoramica sul concetto di filosofia dell'informazione si veda Adriaans 2012/2020.

decisivo per il futuro delle nostre democrazie. Altrettanto pressante e importante è la questione della materializzazione dei valori all'interno delle tecnologie.

Una chiave di lettura per comprendere l'impatto e le conseguenze delle tecnologie digitali in campo etico è la metamorfosi che queste hanno prodotto in una delle categorie fondamentali della modernità: quella di Soggetto, insieme alle caratteristiche che lo definivano come l'unico detentore della moralità. L'idea che il digitale eserciti un potere di re-ontologizzazione e re-epistemologizzazione sul modo in cui percepiamo la realtà e noi stessi (Floridi, 2022, p. 39), e che il suo impatto si estenda alla sfera etica, sociale e politica – malgrado l'urgenza di una riflessione su queste dinamiche è oggi più pressante che mai – non è certo nuova. Norbert Wiener, il padre della cibernetica, prendendo in esame le istanze etiche che avrebbero dovuto condurre la nuova scienza, la *scienza della comunicazione e del controllo nella macchina e nell'animale* verso un *uso umano degli esseri umani*⁴, affermava che:

La società può essere compresa soltanto attraverso lo studio dei messaggi e dei mezzi di comunicazione relativi ad essi; e che nello sviluppo futuro di questi messaggi e mezzi di comunicazione, i messaggi fra l'uomo e le macchine, fra le macchine e l'uomo, e fra macchine e macchine sono destinati ad avere una parte sempre più importante (Wiener, 1963, pp. 23-24).

2. Soggetti frantumati: dall'Io All'I/O

L'impatto rivoluzionario del digitale rappresenta solo l'ultima fase della crisi del Soggetto che abbiamo ereditato dalla modernità filosofica. Le ICT e il pensiero informazionale e computazionale che le accompagnano evidenziano una frattura nelle sue strutture concettuali fondamentali. Il Soggetto moderno non è più monolitico e individuale, unico e identico a sé stesso, come la cartesiana *res cogitans*, ontologicamente ipostatizzata e separata dal mondo, né tantomeno come un "Io" trascendentale kantiano che osserva il mondo.

I germi di questa scissione sono già presenti nel cuore della filosofia moderna, e il digitale non fa che amplificarli. Hobbes, nel *Leviatano*, definisce la caratteristica essenziale del Soggetto, la Ragione, come un calcolo:

⁴ Queste locuzioni non sono casuali. Sono infatti i titoli delle opere maggiori di Wiener, *Cybernetics, or control and communication in the animal and the machine* del 1948 e *The Human Use of Human Beings* del 1950.

Da tutto ciò, possiamo definire (vale a dire determinare) che cosa si vuol dire con questo vocabolo ragione, quando lo calcoliamo tra le facoltà della mente. Infatti, la ragione, in questo senso, non è che il calcolo (cioè l'addizione e la sottrazione) delle conseguenze dei nomi generali su cui c'è accordo per contrassegnare e significare i nostri pensieri; dico contrassegnarli, quando li calcoliamo per noi stessi e significarli, quando dimostriamo o proviamo i nostri calcoli agli altri uomini (Hobbes, 1651, I, Cap. V).

Ragione e pensiero, quindi, non sono altro che un'attività computazionale, un calcolo di simboli. Questa idea espressa da Hobbes nel *Leviatano*⁵ si ritrova, tre secoli dopo, nella teorizzazione dell'Intelligenza Artificiale simbolica – quella che precede la rivoluzione subsimbolica e connessionista del *Deep Learning* – la *Good Old Fashioned Artificial Intelligence* (GOFAI). “Questa è basata su una particolare teoria dell'intelligenza e del pensiero: sostanzialmente l'idea di Hobbes che la ragione sia calcolo” (Haugeland, 1988, p. 107).

Ma chi, o quale istanza, opera queste operazioni? Rispondere a questa domanda comporta l'immediato incontro con un paradosso intrinseco alla concezione computazionale della mente, sia nella modernità filosofica e sia nella GOFAI: il paradosso della ragione meccanica, ovvero la presenza di un *homunculus* intelligente, non puramente meccanico, che manipola i simboli. Questo paradosso implica necessariamente una regressione infinita di *homuncoli* che manipolano simboli (Ivi, pp. 39-43). Non è possibile approfondire ulteriormente qui questo paradosso, ma è importante evidenziare come la teoria computazionale frantumi l'unità del Soggetto.

Il distacco dalla visione tradizionale di soggettività, identità e coscienza, considerate come entità uniche, semplici e completamente trasparenti, si amplifica quando viene preso in esame il paradigma subsimbolico dell'intelligenza artificiale e la correlata teoria della mente, il *Connessionismo*. Questa teoria si fonda su processi sub-cognitivi che operano al di sotto delle strutture logico-simboliche di alto livello, come il linguaggio e la dimensione semantica. I meccanismi che producono gli output delle *Artificial Neural Networks* (ANN), modello esplicativo della mente nelle teorie connessioniste e motore del successo degli algoritmi di *Machine Learning* e *Deep Learning*, sono per loro natura inconoscibili. Questo perché tali processi avvengono all'interno di una *black box* dove la convoluzione dei nodi si sviluppa su più livelli negli strati nascosti delle reti

⁵ È da notare come la definizione di ragione come computazione si presenti per la prima volta nell'opera per eccellenza sul *potere*. Sembra quasi prefigurare la contemporaneità dove il potere computazionale (es. dei giganti del capitalismo digitale) è la base per le altre forme di potere (economico, sociale, politico e così via).

neurale. Pertanto, sia le tecnologie algoritmiche, come l'odierna *Intelligenza Artificiale*, sia la mente umana risultano essere delle scatole nere.

Marvin Minsky, uno dei pionieri dell'AI, parla di "Società della Mente" (Minsky, 1989). Stabilita l'identità tra cervello e mente, Minsky afferma che esiste una pluralità di agenti privi di mente, dalla cui interazione, che una natura "sociale", in un rapporto "talvolta di cooperazione, ma più spesso di conflitto" (Ivi, p. 20) deriva la mente stessa. Non esiste un singolo "trucco magico" che spieghi l'intelligenza (Ivi, p. 308). Di conseguenza, non vi è più un unico principio che possa spiegare e giustificare mente e coscienza; al contrario, sono la diversità e la frammentazione a caratterizzarle.

La disgregazione dell'unità del soggetto moderno "tradizionale" è visibile anche nell'analisi di questa categoria dal punto di vista informativo e nel rapporto con le ICT: il Sé è un Sistema Multi-Agente (MAS) costituito da diverse dimensioni (le "membrane": corporea, cognitiva e cosciente) che interagiscono e si influenzano reciprocamente in un processo poietico (Floridi, 2011). Lo scopo del Sé è semantizzare, ossia trovare schemi di senso nell'ambiente circostante. Il Soggetto si costruisce "narrativamente" nel rapporto con il proprio ambiente, e l'ambiente contemporaneo è sempre più pervaso dalle tecnologie digitali (Ivi). Queste ampliano in modo esponenziale la tipologia e la natura di questo processo autopoietico in relazione ad altri oggetti e al mondo, mettendo in risalto il ruolo decisivo che la mediazione operata dalle ICT riveste e l'importanza, altrettanto decisiva, di un loro design etico.

Prima di approfondire ulteriormente questi temi, è opportuno mettere in luce come, in due dei più celebri *loci* relativi all'Intelligenza Artificiale, si definisca l'intelligenza come azione. Il primo è "l'atto di nascita dell'IA" nel *workshop* di Dartmouth nell'estate del 1956, organizzato da John McCarthy, con la partecipazione di futuri giganti del pensiero computazionale, come Minsky, Rochester, Shannon, Simon e Newell. Il secondo è l'*Imitation Game*, il cosiddetto Test di Turing.

Nella celebre *Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, in cui il termine Intelligenza Artificiale è usato per la prima volta, si stabilisce che: "per l'attuale scopo, il problema dell'intelligenza artificiale è considerato quello di far comportare una macchina in modi che sarebbero definiti intelligenti se fosse un essere umano a comportarsi così"⁶. Il primo vagito dell'IA associa quindi l'intelligenza a un comportamento.

Lo stesso vale per l'*Imitation Game* proposto da Alan Turing nel 1950. Il matematico inglese, dopo aver affermato che la domanda "Possono

⁶ "For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving". McCarthy et al., 1955.

pensare le macchine?” (Turing, 1950) è insensata e andrebbe scartata, propone un gioco per stabilire se le macchine siano intelligenti. Ispirandosi a un popolare gioco di ruolo dell’epoca – in cui tre partecipanti, A (uomo), B (donna) e C (giudice), sono separati e il giudice deve determinare chi sia l’uomo e chi la donna attraverso semplici domande e risposte scritte – Turing riformula la domanda sul pensiero delle macchine in un’ipotesi scientifica verificabile: se una macchina sostituisse uno dei partecipanti e ingannasse il giudice, facendogli credere di essere un essere umano, allora alla macchina andrebbe attribuita l’intelligenza.

Anche in questo caso, l’intelligenza, o meglio, il giudizio sull’intelligenza, è legato all’azione, al comportamento. Nell’azione, le differenze tra uomo e macchina sfumano, non trovando più un ostacolo epistemologico nella concezione sostanziale e ipostatizzata della soggettività intelligente. Se l’intelligenza si dissolve nell’agire, e questo è mediato dagli artefatti tecnologici, è necessario interrogarsi su come questi artefatti modifichino l’*agency* morale.

L’analisi computazionale e informazionale della soggettività evidenzia come, più che un’irriducibile conflittualità tra “Uomo” e “Macchina” – entrambi concetti filosoficamente problematici – le ICT, il digitale e l’AI debbano essere compresi, con distinzioni nette che vanno sfumando, in relazione alla capacità di agire. È qui, nel legame con l’azione – che le tecnologie digitali amplificano, mediano e concretizzano – che va ricercata la dimensione etica delle nuove tecnologie. Queste hanno contribuito a modificare l’*agency* e la responsabilità morale tipica della modernità e pongono il problema di una loro implementazione eticamente sostenibile.

3. Trasformazioni dell’*agency* e della responsabilità: una genealogia sociotecnica

Le riflessioni precedentemente esposte sulla trasformazione, sia ontologica che epistemologica, della natura del soggetto, causata dalla rivoluzione digitale, sollevano un ulteriore tema. Se le ICT e l’AI rappresentano oggi una riserva della nostra capacità di agire, diventa essenziale comprendere l’impatto che tali tecnologie hanno avuto sul concetto di *agency* e come esse, già dalla loro progettazione, mediano e trasformano il nostro agire morale.

Se l’introduzione delle ICT ha messo in discussione la concezione del Soggetto ereditata dalla modernità filosofica, allora le categorie etiche tradizionali, che attribuivano al Soggetto “moderno” il monopolio della moralità, non sono più sufficienti per giudicare l’azione mediata dagli artefatti digitali. Questi artefatti, come abbiamo già osservato, possono non solo essere agenti autonomi, ma anche costituire il mezzo attraverso

cui si costruisce narrativamente la nostra identità. Di conseguenza, diventa imprescindibile considerare nuovi fattori che rendono insostenibile il modello tradizionale di relazione lineare tra utente e oggetto nella produzione di atti moralmente qualificabili. V. Flusser offre una sintesi efficace e originale della metamorfosi in corso.

Negli anni '90 del secolo scorso, Flusser osservava come fosse diventato estremamente arduo attribuire la responsabilità morale in una società in cui le tecnologie sono sempre più pervasive. Secondo il pensatore ceco-brasiliano, questo aumento di complessità era dovuto a tre fattori principali:

1) La perdita di credibilità delle autorità tradizionali (religiose, politiche, morali) nel definire norme sociali, particolarmente nel contesto di una produzione industriale complessa;

2) La crescente complessità della produzione industriale e del design, che comportano una quantità di informazioni tale da eccedere la capacità cognitiva individuale, richiedendo una cooperazione organizzativa nei processi di progettazione e implementazione. L'azione, spesso svolta da "team" che combinano componenti umane e artificiali, non può pertanto essere attribuita a un singolo autore: "Per questo motivo, nessuna persona può essere ritenuta responsabile di un prodotto";

3) In passato, la responsabilità morale ricadeva solo sull'utilizzatore di un prodotto. "Se una persona pugnava un'altra con un coltello, ne era l'unico responsabile, non il designer del coltello". In questo scenario, la produzione di coltelli era considerata un'attività pre-etica, priva di valutazioni morali. Tuttavia, oggi, con l'automazione dei processi produttivi, ci si potrebbe chiedere se, ad esempio, siano i robot o le macchine a essere responsabili dell'uso dei prodotti. Dal momento che questo è un assurdo, a chi deve essere attribuita la responsabilità morale?

Alla persona che ha costruito il robot, a quella che ha realizzato il coltello, oppure a chi ha programmato il funzionamento del robot? Si potrebbe forse attribuire la responsabilità a un errore di progettazione, di programmazione o di produzione? E perché non imputare la responsabilità morale all'intero settore industriale che ha prodotto il robot? O magari all'intero complesso industriale, e, in ultima analisi, al sistema di cui tale complesso fa parte? (Flusser, 2003, pp. 63-65).

Queste domande evidenziano come il quadro della responsabilità morale sia cambiato irreversibilmente. Questi problemi non sono contingenti o legati esclusivamente allo sviluppo industriale del XX secolo, ma sono strettamente legati al ruolo degli artefatti tecnologici nel nostro agire e al loro design. Un esempio particolarmente drammatico, utilizzato da Flusser per illustrare il ruolo della progettazione/produzione nell'*agency* morale, è emerso durante il processo di Norimberga. In una lettera scritta

da un industriale tedesco a un ufficiale nazista, l'industriale si scusava per aver progettato male le camere a gas che gli erano state commissionate: “invece di uccidere migliaia di persone per volta, ne venivano uccise solo centinaia” (Ivi, p. 66).

L'episodio è emblematico di come “a) non esistano più norme applicabili alla produzione industriale; b) non ci sia più un singolo autore di un crimine; e c) la responsabilità sia così diluita che di fatto ci troviamo in una situazione di totale irresponsabilità rispetto agli atti derivanti dalla produzione industriale” (Ibidem).

A questi tre punti possiamo aggiungere una considerazione ulteriore: il ruolo di un artefatto, come nell'esempio citato da Flusser, nel determinare un risultato moralmente giudicabile. Non è più possibile considerare la relazione tra oggetto e utente come lineare, senza tener conto del designer e del processo di progettazione. I designer, dunque, non possono eludere una riflessione etica sul loro lavoro, pena il totale dissolvimento della responsabilità morale individuale. L'episodio sopra menzionato mette drammaticamente in evidenza la funzione di mediazione degli artefatti tecnologici nella sfera morale.

4. Mediazione tecnologica e metamorfosi fenomenologiche

Un contributo decisivo all'analisi del concetto di mediazione tecnologica, dal punto della *postfenomenologia* – sulla scia del concetto di *script* di Latour (1992) –, è stato quello di P. P. Verbeek. Il filosofo olandese sostiene che la mediazione tecnologica è costitutiva dell'azione degli utilizzatori, con le tecnologie che plasmano le azioni dell'utente oltre la sfera della mera funzionalità. Il design iscrive modalità d'uso specifiche nella tecnologia, con un carattere interazionale. “Le tecnologie hanno “intenzioni,” non sono strumenti neutrali ma giocano un ruolo attivo nella relazione tra gli esseri umani e il loro mondo” (Verbeek, 2006)⁷. Gli artefatti tecnologici non solo influenzano il modo in cui gli esseri umani agiscono, ma anche il modo in cui organizzano la loro vita sociale e ambientale.

Le riflessioni di Verbeek implicano una responsabilità condivisa tra utenti e designer per le conseguenze etiche delle azioni mediate dalla tecnologia. Responsabilità che si articola in tre modalità:

1. Responsabilità dei progettisti;
2. Responsabilità degli utilizzatori;
3. Consapevolezza etica condivisa.

⁷ “Technologies have ‘intentions’, they are not neutral instruments but play an active role in the relationship between humans and their world”.

La mediazione tecnologica deve essere integrata pienamente nella riflessione etica. “Gli artefatti tecnologici non sono intermediari neutrali ma plasmano attivamente il modo in cui le persone sono nel mondo: le loro percezioni e azioni, l’esperienza e l’esistenza”⁸ (Ivi). Occorre riconoscere che le tecnologie non sono semplici strumenti passivi, ma coinvolgono attivamente gli esseri umani e altri attori non umani in assemblaggi di azione che influenzano il nostro modo di vivere. Sia i progettisti che gli utenti devono essere consapevoli delle implicazioni etiche connesse alle tecnologie, adottando un approccio critico e olistico che consideri le complesse interazioni tra esseri umani, tecnologie e ambiente. In questo senso, attraverso la mediazione tecnologica, il design realizza materialmente la moralità, poiché offre risposte mediate alla domanda su come si debba agire, delineando la responsabilità morale specifica dei designer⁹. Il concetto di *agency* morale si evolve: non è più limitato al solo essere umano (o al Soggetto tradizionale), ma include anche l’oggetto, la sua non neutralità etica e la storia della sua progettazione. Queste istanze si comprendono meglio

L’ambiente informazionale, perennemente connesso e potenziato dalla globalizzazione, dal digitale e dalle opportunità offerte dalla mediazione tecnologica, dimostra chiaramente come le categorie etiche tradizionali non siano più sufficienti, ma necessitino di un ampliamento. Van den Hoven sottolinea come, insieme al mutamento del contesto sociale in cui si svolge l’azione morale, sia cambiata anche quella che S. Scheffer definisce la fenomenologia dell’*agency* ispirata al senso comune.

La concezione tradizionale della moralità prevedeva:

1. la priorità dei doveri negativi (evitare di danneggiare gli altri) rispetto ai doveri positivi di fare il bene o aiutare gli altri nel perseguimento dei loro obiettivi;
2. una divisione del lavoro morale, in cui gli obblighi verso le persone più vicine sono più forti rispetto a quelli verso individui lontani, sia nello spazio che nella società;

⁸ “Technological artifacts are not neutral intermediaries but actively shape people’s being in the world: their perceptions and actions, experience, and existence”

⁹ La responsabilità del designer è sempre più chiamata in causa, via via che la capacità persuasiva degli artefatti diviene maggiore a causa dello sviluppo tecnologico. Il testo chiave per comprendere le tecnologie della persuasione è Fogg, 2017. Fogg, specialmente grazie all’influenza del suo *Behaviour Design Lab* a Stanford, è “l’eminenza grigia” che sta dietro molte delle testimonianze di protagonisti delle industrie hi-tech contenute nel documentario di Netflix, *The Social Dilemma* del 2020. Per quanto riguarda la responsabilità morale dei designer in relazione ad azioni intenzionali o non-intenzionali ed effetti prevedibili o non prevedibili, si rimanda a Berdichevsky, Neuenschwander, 1999.

3. il primato dell'atto sull'omissione, con l'atto considerato sempre come autonomo, mentre l'omissione è vista come contestuale;
4. la preminenza degli effetti immediati delle nostre azioni rispetto a quelli lontani. Inoltre, la percezione della nostra influenza causale diminuisce proporzionalmente alla distanza spazio-temporale degli effetti. La nostra *agency* individuale, infine, viene percepita in modo attenuato quando agiamo in contesti sociali complessi o in collaborazione con altri (Van den Hoven, 2017).

Questa fenomenologia dell'*agency* non è più in grado di rispondere adeguatamente alle sfide del mondo contemporaneo. Come, nel paragrafo precedente le analisi di Flusser hanno mostrato, non è più possibile considerare una semplice relazione lineare tra agente/utilizzatore e oggetti. L'agente morale individuale, modellato da questa visione, oggi si trova subordinato a strutture globali che dirigono, interconnettono e influenzano profondamente la sua azione morale. Un esempio emblematico di ciò è la facilità con cui è possibile compiere azioni eticamente disastrose utilizzando semplicemente un computer connesso a Internet.

Proseguendo con l'analisi della tesi di Scheffer, Van den Hoven osserva che oggi “1) il contributo degli individui ai risultati complessivi è limitato; 2) il loro controllo è ridotto; 3) i processi in cui sono coinvolti sono pervasivi e inevitabili; e 4) è molto difficile ottenere informazioni accurate, affidabili e tempestive su tali processi e i loro effetti”¹⁰ (Ivi). Questa situazione, come descritto da Scheffer, evidenzia chiaramente che la nostra concezione della moralità, così come espressa dalla fenomenologia del senso comune, necessita di essere ampliata. Tuttavia, per Van den Hoven, questo non è sufficiente a spiegare un fattore ulteriore:

Molte delle azioni che compiamo nel ventunesimo secolo non le compiamo direttamente l'uno verso l'altro, ma quasi sempre tramite la tecnologia e attraverso un milieu tecnologico che è stato progettato e che è emerso nel corso dei secoli a causa di innumerevoli contributi da parte di altri separati da noi nel tempo e nello spazio (Ivi).¹¹

¹⁰ “The subsumption of individual agents in complex global interconnected structures has a number of morally relevant features according to Scheffer: (1) the contribution of individuals to overall results is limited; (2) their control is limited; (3) the processes in which they are subsumed are pervasive and inescapable; and (4) it is very difficult to get accurate, reliable and timely information about these processes and their effects”

¹¹ “Much of what we do in the twenty-first century we do not do directly to each other, but almost always by way of technology and via a technological milieu that was designed, and that came about over the course of centuries as a result of myriad contributions by others separated from us by time and place”

Il Soggetto della contemporaneità, influenzato dall'impatto delle ICT, si costruisce narrativamente attraverso l'interazione con un ambiente sempre più permeato da artefatti tecnologici e informativi. Il ruolo del design nella considerazione morale implica che la riflessione etica debba essere accompagnata da una genealogia del design, capace di spiegare la stratificazione delle scelte progettuali: "se Nietzsche ha tracciato la genealogia della morale, la svolta progettuale ha tracciato la genealogia delle scelte progettuali" (Umbrello, 2023). La trasformazione dell'*agency* e del soggetto agente rende obsoleta la tradizionale relazione diretta tra agente e oggetto. La metamorfosi dell'ambiente tecnologico, unita alla pluralità di stakeholder coinvolti (designer, utenti, pazienti indiretti, agenti e pazienti artificiali o ibridi), riformula la fenomenologia morale, chiarendo il cambiamento qualitativo e quantitativo dell'*agency* rispetto alla moralità dei soggetti tradizionali.

Anche alla luce di queste considerazioni sul Soggetto, diventa allora evidente l'importanza cruciale della progettazione e della sua storia. Nella riflessione morale, è imprescindibile prendere in esame "1) le storie di progettazione dei sistemi e dell'ambiente tecnologico in cui si realizza l'*agency*; e 2) le scelte e il ruolo degli agenti di progettazione nella formazione delle architetture di scelta" (Ivi).

In qualunque ambiente informativo – sia esso quello operativo diretto dell'utente o quello integrato in un sistema di controllo o di utilizzo – le scelte di design risultano decisive. Un operatore nella sala di controllo del traffico aereo, un trader di borsa, un chirurgo in una moderna sala operatoria o un tecnico in una centrale nucleare hanno precise responsabilità e doveri morali. Tuttavia questi agenti "non possono svolgere la loro azione [...] senza un ambiente di lavoro tecnologico adeguatamente progettato. Se il software dei loro sistemi contiene bug, se i modelli utilizzati si basano su presupposti errati o se le informazioni sono imprecise, le decisioni e le azioni umane erediteranno questi difetti"¹² (Ivi).

In questo contesto, gli agenti, umani e non, interagiscono con artefatti e ambienti informativi generati da processi di progettazione e implementazione. Il *design turn* dell'etica diventa così il mezzo per dare forma ai valori in un'epoca in cui i confini tradizionali tra le categorie stabilite si fanno sempre più sfumati. La capacità di coniugare sviluppo tecnologico e dimensione etica emerge quindi come una sfida cruciale per le società

¹² "Think of an operator in the control room of a nuclear power plant, an air traffic control leader, a surgeon in a modern operating theatre, or a trader in a trading room. All of these professionals have duties and responsibilities, and they cannot possibly do what they are expected to do — and what everyone justifiably expects them to do — without a properly designed technological work environment. If the software of their systems contains bugs, the models used are based on incorrect assumptions, or the information is inaccurate, then human decisions and actions will inevitably inherit these flaws"

democratiche, chiamate a trovare un equilibrio tra innovazione e rispetto dei principi morali fondamentali.

5. Etica e design tecnologico

La rivoluzione digitale ha messo in luce come il nostro ambiente sia sempre più di natura informazionale, caratterizzato da livelli di complessità crescenti. Viviamo e agiamo all'interno dell'*Infosfera*¹³, un'epoca dominata dai *Big Data* e dalla proliferazione di *stakeholder*, che arriva a comprendere “specie non umane, superorganismi, la Terra stessa, i social robot”¹⁴ (Friedman, Hendry, 2019, p. 27). In questo orizzonte, la sfera legislativa, sia a livello nazionale che sovranazionale, ha progressivamente posto al centro dell'attenzione la dimensione etica delle nuove tecnologie, con particolare riferimento alla loro progettazione e implementazione. Un chiaro esempio di questo approccio è rappresentato dal *Regolamento Generale sulla Protezione dei Dati* (GDPR) dell'Unione Europea, entrato in vigore nel 2018, che introduce concetti fondamentali relativi alla gestione e protezione dei dati personali. Tra le novità principali sancite dal GDPR, e recepite nelle legislazioni nazionali, emergono i concetti di *Privacy by Design* e *Privacy by Default*, definiti rispettivamente nell'art. 25 e nel considerando 78 del Regolamento.

Il principio di *Privacy by Design* sottolinea come le istanze etiche, come la protezione della privacy, debbano essere incorporate direttamente nei sistemi mediante il loro design. Ciò significa che la protezione dei dati personali deve essere integrata sin dalle prime fasi di progettazione di sistemi, processi, prodotti o servizi che trattano tali dati, rendendo la privacy una parte intrinseca dell'architettura tecnica e concettuale delle tecnologie stesse.

Il concetto di *Privacy by Default*, complementare a quello di *Privacy by Design*, stabilisce che i sistemi devono essere configurati in modo da garantire automaticamente il massimo livello di protezione dei dati personali, senza che l'utente debba intervenire per modificarne le impostazioni. In questo modo, le impostazioni predefinite minimizzano la raccolta e l'uso dei dati, rafforzando il rispetto dei diritti degli utenti.

¹³ “L'infosfera indica l'intero ambiente informazionale costituito da tutti gli enti informazionali, le loro proprietà, interazioni, processi e reciproche relazioni”. Il concetto può essere esteso sino a comprendere l'intero reale. L. Floridi, 2017, p. 44.

¹⁴ “Yet the technologies we design and build reach far beyond human beings to implicate other nonhuman entities. Here are four for consideration: nonhuman species, superorganisms, the Earth, and social robots”

Quanto sancito dal GDPR riflette la crescente consapevolezza normativa della necessità di affrontare le sfide etiche poste dall'uso delle tecnologie digitali e degli algoritmi, assicurando che istanze etiche fondamentali, come la privacy¹⁵, siano tutelate fin dalla fase di progettazione delle tecnologie stesse. In questo contesto, si può riconoscere una formale consacrazione del *design turn* nell'etica applicata, in cui il processo di progettazione diventa centrale per garantire il rispetto dei principi etici (Van den Hoven, 2017). Sebbene non priva di criticità in alcune metodologie, come il *Value Sensitive Design* (VSD)¹⁶ attraverso cui tenta di concretizzarsi, la svolta progettuale dell'etica, ovvero l'incarnazione dei valori negli artefatti, può rappresentare una via verso un'innovazione sostenibile. Insieme ad approcci come l'*Artificial Intelligence for Social Good* (AI4SG)¹⁷, essa offre

¹⁵ Per una panoramica sull'implementazione della privacy tecnologica, che sfata i miti sulla protezione dei dati attraverso algoritmi, si veda Hoepman, 2021. Per una trattazione snella ed efficace delle strategie di progettazione orientate alla privacy, si rimanda a Hoepman, 2022.

¹⁶ Il VSD è un approccio teorico alla progettazione tecnologica che incorpora sistematicamente i valori umani nel processo di design, sviluppato da Batya Friedman e colleghi negli anni '90. Il VSD adotta una metodologia tripartita, basata su indagini concettuali, empiriche e tecniche, e riconosce l'importanza di valori morali come giustizia, benessere e dignità (Friedman et al., 2008; Friedman, Hendry, 2019). Tuttavia, tra le principali criticità emerge la mancanza di una chiara definizione di "valore", oltre al rischio di oscillare tra relativismo e universalismo. In particolare, viene imputata al VSD la tendenza all'ipostatizzazione valoriale, ovvero la tendenza a reificare un insieme ristretto di valori che potrebbe non adattarsi ai contesti locali, ostacolando la scoperta di nuovi valori emergenti. Il VSD tenderebbe, inoltre, alla "fallacia naturalistica", dove si confonde ciò che è con ciò che dovrebbe essere. Questo problema sottolinea la difficoltà di bilanciare l'universalità dei valori con la loro adattabilità ai diversi contesti socioculturali e progettuali, una questione ancora aperta nella metodologia del VSD. Per le critiche al VSD, si veda Albrechtslund, 2007; Le Dantec et al., 2009; Mander-Huits, 2011. Intorno al rapporto VSD-AI4SG si veda Umbrello, 2023. Gli oggetti, come evidenziato nell'analisi, non sono neutri; al contrario, essi operano mediando l'azione e esercitano una forma di influenza, più o meno diretta, che può manifestarsi sotto forma di spinta, di *nudging*: "Qualsiasi aspetto dell'architettura delle scelte che altera il comportamento delle persone in modo prevedibile senza vietare alcuna opzione o cambiare significativamente i loro incentivi economici". Thaler & Sunstein 2009, p.6, trad. mia. L'incarnazione di valori nella materialità degli oggetti, l'*ethics by design*, può essere vista come una forma di *nudging* ontologico, che viola l'autonomia decisionale del soggetto che non può non agire come l'oggetto impone. A questa forma in cui rientra il VSD L. Floridi oppone il *pro-ethical design*, una forma di *nudging* informazionale, modifica la natura delle informazioni a cui le persone sono esposte, influenzando così le decisioni senza alterare le opzioni fisiche disponibili, come più rispettoso dell'autonomia dell'agente. La forma dell'*ethics by design*, via scelta dagli organismi legislativi dell'Unione Europea per proteggere beni etici in relazione all'implementazione tecnologica, può trovare da queste riflessioni spunti per una progressiva riduzione di potenziali criticità etiche. Floridi, 2016.

¹⁷ L'AI4SG è il "design, lo sviluppo e l'implementazione di sistemi di IA in modo da (i) prevenire, mitigare o risolvere i problemi che incidono negativamente sulla vita umana o/e sul benessere del mondo naturale e/o (ii) consentire sviluppi preferibili dal punto di vista sociale e/o sostenibili dal punto di vista ambientale.". Floridi, 2022, p. 227.

una prospettiva promettente per coniugare progresso tecnologico e rispetto dei principi etici.

L'aver preso in esame questi snodi critici dell'impatto dello sviluppo tecnologico delle ICT nell'orizzonte socio-tecnico del digitale, può essere utile per offrire spunti affinché moralità e valori trovino maggiore tutela nell'implementazione tecnologica e nella produzione legislativa, incarnandosi in quegli artefatti che influenzano profondamente le nostre azioni eticamente significative, permettendone un uso rispettoso delle prerogative umane.

Bibliografia

- Albrechtslund, A.
2007 *Ethics and technology design*. *Ethics and Information Technology*, 9, 63-72. <https://doi.org/10.1007/s10676-006-9129-8>
- Allen, C., Varner, G., Zinser, J.
2000 *Prolegomena to any future artificial moral agent*, *Journal of Experimental & Theoretical Artificial Intelligence*, 12(3), 251-261. DOI:10.1080/09528130050111428
- Allo, P. (guest editor)
2010 *Special Issue: Luciano Floridi and the Philosophy of Information*, 41(3), *Metaphilosophy*, Wiley. <https://doi.org/10.1111/j.1467-9973.2010.01642.x>
- Bender, E. M., Gebru, T., McMillan-Major, A., Shmitchell, S.
2021 *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623. <https://doi.org/10.1145/3442188.3445922>
- Berdichevsky, D., & Neuenschwander, E.
1999 *Toward an Ethics of Persuasive Technology*. *Communications of the ACM*, 42(5), pp. 51-58.
- Dennett, D.
1997 *When Hal Kills, Who's to Blame? Computer Ethics*, in D. Stork (a cura di), *Hal's Legacy: 2001's Computer as Dream and Reality*, MIT Press, Cambridge, MA, pp. 351-365.
- Floridi, L.
2008 *The method of levels of abstraction*, *Minds and Machines*, 18(3), 303-329. <https://doi.org/10.1007/s11023-008-9113-7>
- 2010 *Information Ethics*, in *The Cambridge Handbook of Information and Computer Ethics*, Edited by Luciano Floridi, Cambridge University Press, Cambridge.
- 2011 *The informational nature of personal identity*, *Minds and Machines*, 21(4), 549-566. DOI:10.1007/s11023-011-9259-6
- 2016 *Tolerant Paternalism: Pro-ethical Design as a Resolution of the Dilemma of Toleration*, *Science and Engineering Ethics*, 22(6), 1669-1688. <https://doi.org/10.1007/s11023-011-9259-6>
- 2017 *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Raffaello Cortina, Milano.

- 2020 *Pensare l'infosfera, la filosofia come design concettuale*, Raffaello Cortina, Milano.
- 2022 *Etica dell'Intelligenza Artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina, Milano.
- Floridi, L., Sanders, J. W.
- 2004 *On the morality of artificial agents*, *Minds and Machines*, 14(3), 349-379. <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>
- Flusser, V.
- 2003 *Filosofia del design*, Bruno Mondadori, Milano.
- Friedman, B., Hendry, D. G.
- 2019 *Value Sensitive Design: Shaping Technology with Moral Imagination*, MIT Press Cambridge MA.
- Friedman, B., Kahn, P. H. Jr., Borning, A.
- 2008 *Value Sensitive Design and Information Systems*, in K.E. Himma, H.T. Tavani (a cura di), *The Handbook of Information and Computer Ethics*, John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 69-101.
- Gleick, J.
- 2012 *L'informazione: Una storia. Una teoria. Un diluvio*, Feltrinelli Milano.
- Haugeland, J.
- 1988 *Intelligenza Artificiale. Il significato di un'idea*, Bollati Boringhieri Torino.
- Hobbes, T.
- 1651 *Leviathan, or The Matter, Forme and Power of a Commonwealth Ecclesiastical and Civil*; trad. it, *Il Leviatano*; BUR, Milano 2011.
- Hoepman, J.-H.
- 2021 *Privacy Is Hard and Seven Other Myths: Achieving Privacy through Careful Design*. Cambridge, MA: The MIT Press.
- 2022 *Privacy Design Strategies (The Little Blue Book)*. <https://creativecommons.org/licenses/by-nc/4.0/>.
- LATOURE, B.
- 1992 *Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts*, in W. E. Bijker, J. Law (a cura di), *Shaping Technology/Building Society: Studies in Sociotechnical Change*, MIT Press, Cambridge, MA, 225-258.
- Le Dantec, C. A., Poole, E. S., & Wyche, S. P.
- 2009 *Values as lived experience: Evolving value sensitive design in support of value discovery*. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1141-1150. New York, NY: ACM. <https://doi.org/10.1145/1518701.1518875>
- Manders-Huits, N.
- 2011 *What values in design? The challenge of incorporating moral values into design*. *Science and Engineering Ethics*, 17(2), 271-287. <https://doi.org/10.1007/s11948-010-9198-2>
- McCarthy, J., Minsky, M. L., Rochester, N., Shannon, C. E.
- 1955 *A proposal for the Dartmouth summer research project on artificial intelligence*, *AI Magazine*, 27(4).
- Minsky, M.
- 1989 *La società della mente*, Adelphi Milano.
- Shannon, C. E.
- 1948 *A Mathematical Theory of Communication*, *Bell System Technical Journal*, 27(3), 379-423.

- Thaler, R. H., Sunstein, C. R.
2008 *Nudge: Improving decisions about health, wealth, and happiness*, Yale University Press, New Haven CT.
- Turing, A. M.
1950 *Computing Machinery and Intelligence*, *Mind*. Vol. 59, No. 236 (Oct., 1950), pp. 433-460; trad. it. in *Intelligenza Meccanica*, Bollati Boringhieri, Torino 1994.
- Umbrello, S.
2023 *Oggetti buoni. Per una tecnologia sensibile ai valori*, Fandango Roma.
- Van Den Hoven, J.
2017 *The Design Turn in Applied Ethics*, in J. van den Hoven, S. Miller, T. Pogge (a cura di), *Designing in Ethics*, Cambridge University Press, Cambridge, 11-31. DOI:10.1017/9780511844317.002
- Verbeek, P. P.
2006 *Materializing Morality: Design Ethics and Technological Mediation*, *Science, Technology, & Human Values*, 31(3), 361-380. DOI:10.1177/0162243905285847
- Wiener, N.
1948 *Cybernetics; or control and communication in the animal and the machine*. John Wiley & Sons New York.
- 1950 *The Human Use of Human Beings: Cybernetics and Society*, Houghton Mifflin, Boston. trad. it. *Introduzione alla cibernetica – L'uso umano degli esseri umani*, Einaudi Torino, 1953

Giorgia Redi*

RoboCare. Anziani più indipendenti o più invisibili?

Abstract

RoboCare. More independent or more invisible elders?

In recent decades, the use of robots has increased dramatically, and it is easily foreseeable that they will become even more a part of our lives in the coming years. In this paper, we will turn our attention to the role of *Socially Assistive Robots* (SAR), designed to alleviate loneliness and provide a valuable aid in the activities of caring and caring for the elderly for both caregivers and their “users”. Living in a country characterized by an increasingly low birth rate and an increasingly ageing society, there will be in the future fewer and fewer young people willing to care full-time for sick, elderly or disabled people. In view of such predictions, we consider it necessary to reflect critically about the use of robots in the care sector, which are built to become good caregivers.

Conceived to coexist with humans and to perform functions traditionally attributed to human beings, social robots induce us to rethink our way of conceiving and living a relationship, finding ourselves before an Other who is not classically a “body” placed before us. Finally, we will be forced to rethink the impact of a human-machine relationship, which will become a caring relationship, where the caregiver is a so-called “substitute” and the care recipient a flesh-and-blood human being, more fragile and vulnerable. This reflection will also be an opportunity to question the need to sustain an “exploratory” moral practice, given the heterogeneity of the agents – human and robotic – involved, soon to change the future of medicine, the social-health sector and the more traditional care relationships.

Keywords

Ageing; Loneliness; Social Robots; Caring Relationships; Ethics of Care

* Università di Napoli “Federico II”.



1. Introduzione

Kokoduschi. Letteralmente: “morte per solitudine”. Un fenomeno invisibile che in Giappone comporta la morte di una media di 68.000 anziani all’anno, a fronte del progressivo invecchiamento della popolazione. Secondo i dati ANSA riportati dall’Agenzia nazionale di polizia giapponese¹, nonostante gli sforzi del governo per combattere l’isolamento sociale cui spesso conseguono tragici epiloghi, la morte solitaria riguarderebbe per lo più uomini e donne con un’età superiore ai 65 anni. Tuttavia, se il Giappone si distingue come nazione avente la popolazione più anziana del mondo, non è necessario giungere sino all’Oceano Pacifico per scorgere risultati simili. L’Italia, infatti, si classifica al secondo posto per longevità su scala globale, continuando a confermarsi uno dei Paesi più “vecchi”² dell’UE e non solo.

Non è un caso che, soprattutto nelle società e nei sistemi di *welfare* occidentali, gli ultimi venti anni siano stati caratterizzati dalla presenza di due fenomeni contrapposti che continueranno a permeare gli anni a venire. Al sorprendente aumento dell’aspettativa di vita della popolazione è conseguita una maggiore incidenza e diffusione delle malattie neurodegenerative, divenute in alcuni contesti vere e proprie epidemie (ne è un esempio la demenza di Alzheimer³). Malattie che interessano tutte le fasi della vita benché ad essere colpiti siano per lo più gli anziani, come rilevato da una delle recenti attività dell’ISS⁴. Dunque, la nostra è una popolazione longeva ma allo stesso tempo più malata e bisognosa di cure.

Emergono allora due situazioni alquanto critiche: la sostenibilità della sanità pubblica, pressata dal costante invecchiamento della popolazione e dalla carenza del personale sanitario, e la massiccia presenza di persone anziane che vivono da sole, spesso non autosufficienti e richiedenti un sostegno costante e continuativo. Quest’ultimo è l’aspetto che qui vorremo prendere in considerazione, venutosi a delineare come un’emergenza

¹ Redazione ANSA (a cura di), *In Giappone ogni anno 68000 anziani muoiono in solitudine, mentre accelera il processo di invecchiamento della popolazione*, in “ANSA.it”, 15/05/2024, ultimo accesso 11/11/2024, https://www.ansa.it/sito/notizie/mondo/2024/05/15/in-giappone-ogni-anno-68mila-anziani-muoiono-in-solitudine_72dc21d4-9267-45c5-a10d-4d2841e62da2.html.

² M. Chelli, *Rapporto annuale 2024. La situazione del Paese*, ISTAT, Roma 2024, ultimo accesso 11/11/2024, <https://www.istat.it/wp-content/uploads/2024/05/Sintesi-Rapporto-Annuale-2024.pdf>.

³ Ministero della Salute, *Dati epidemiologici. Demenze*, ultima modifica 20/09/2023, ultimo accesso 12/11/2024, <https://www.pnrr.salute.gov.it/portale/demenze/dettaglioContenutiDemenze.jsp?lingua=italiano&id=2402&area=demenze>.

⁴ Istituto Superiore della Sanità, *Malattie croniche e invecchiamento in salute*, ultima modifica 05/01/2022, ultimo accesso 10/11/2024, <https://www.iss.it/malattie-croniche-invecchiamento-in-salute>.

silenziosa e passata inosservata tanto in Italia quanto a livello globale. “La solitudine è il nuovo fumo di sigaretta”⁵, uno slogan di impatto coniato da Vivek H. Murthy e corroborato da recenti evidenze scientifiche che associano l’isolamento sociale ad un’accelerazione dell’invecchiamento fisiologico oltre che ad un maggiore rischio di sviluppare un rapido declino cognitivo, e più genericamente di compromettere il benessere mentale.

Se ancora ad oggi il sistema sanitario delega principalmente ai *caregiver informali* – causa la pregressa carenza di operatori sanitari e i costi drammatici dell’assistenza istituzionale a lungo termine nelle case di cura o in altre strutture sanitarie – compiti relativi al supporto e alla cura quotidiana dei propri cari, anche l’assistenza informale continua a subire una drastica riduzione. In ciò, si può scorgere una maggiore indipendenza delle giovani generazioni rispetto ad un tempo e una maggiore partecipazione delle donne al mondo del lavoro (tradizionalmente deputate al lavoro di cura). Pertanto, i membri della famiglia, in molti casi, vivrebbero lontano dai propri cari e i nuclei familiari tenderebbero ad essere più piccoli oltre che fagocitati da quell’impostazione capitalista ed efficientista che oramai domina le nostre vite. La *cronofagia*, cioè la sensazione di venire “divorati” dal tempo, ci conduce così a non dedicare più sufficienti attenzioni e cure ai nostri anziani, sia che presentino un invecchiamento fisiologico o che ne manifestino uno patologico.

Volendo individuare nuove modalità di sostegno e di supporto a coloro che prestano assistenza formale e informale alle persone anziane e alle loro famiglie, proponiamo qui una riflessione in merito all’utilizzo dell’intelligenza artificiale e della robotica⁶ nel settore socioassistenziale, settore storicamente affidato all’essere umano. Dopo avere fotografato alcuni degli esiti odierni, frutto della collaborazione tra l’ingegno umano e l’utilizzo dell’intelligenza artificiale nella progettazione e nella realizzazione di agenti artificiali da impiegare in ambito assistenziale, il nostro obiettivo non sarà valutarne l’efficacia nell’erogazione di assistenza bensì quello di indagare nuove modalità relazionali ed emozionali tra esseri umani ed agenti robotici. Analizzeremo, in breve, alcune delle implicazioni morali derivanti da tali trasformazioni demografiche e sociali, per

⁵ Cfr. V.H. Murthy, *Letter from the Surgeon General*, in Office of the Surgeon General (a cura di), *Our Epidemic of Loneliness and Isolation. The U.S. Surgeon General’s Advisory on the Healing Effects of Social Connection and Community*, US Department of Health and Human Services, Washington 2023, pp. 4-5, ultimo accesso 13/11/2024, <https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf>.

⁶ Riguardo il “cosa sia un robot”, non esiste oggi una plausibile definizione in grado di mettere d’accordo tutti i roboticisti. Questi ultimi sono soliti distinguere i robot in base alle sembianze (umanoidi, non umanoidi, pet robot etc.), alle funzioni oppure allo scopo per il quale vengono commercializzati. Cfr. P. Lin, K. Abney, G.A. Bekey, *Robotic Ethics. The Ethical and Social Implications of Robotics*, MIT Press, Cambridge 2012, p. 17.

avanzare una disamina ed un ripensamento di due termini fondamentali, quali “relazione” e “cura”, nonché del loro ruolo in una società ipercognitivista⁷. Potrà la robotica, un giorno, contribuire ad affrontare tutte queste nuove sfide legate all’invecchiamento della popolazione globale? I mutamenti sociali in atto troveranno soluzione nella possibilità di affidare il lavoro di cura all’apparato tecnologico dell’automazione?

2. Le critiche etiche esistenti sull’Altro robotico

C’è qualcosa di inquietante e promettente al tempo stesso nella prospettiva di impiegare i robot cosiddetti “sociali”⁸ nell’assistenza agli anziani, che sembra consistere in una possibile soluzione all’erosione di risorse temporali, umane ed economiche disposte ad impegnarsi nell’attività di cura. Asserendo che l’indipendenza e l’emancipazione sono oggi condizioni auspicabili e desiderabili, così come il vantaggio di poter trascorrere la vecchiaia nella propria abitazione privata, ecco che l’utilizzo della robotica sociale in questo contesto potrebbe portare a novità rilevanti capaci di rivoluzionare l’attuale gamma di possibilità.

Sono rinvenibili numerosi modelli di robot sociali “amichevoli” impiegati nei centri di cura e negli ospedali (principalmente in Giappone, ma anche negli Stati Uniti e in alcuni Paesi dell’UE, dove non mancano sperimentazioni in Italia⁹) con svariate specifiche funzioni di supporto alla persona: dalla comunicazione verbale al lavoro di cura cognitiva ed emotiva, dal monitoraggio dei parametri clinici alla riabilitazione, dal ricordare al proprio utente l’assunzione di farmaci alla promozione dell’interazione sociale, etc. Tra questi, alcuni hanno ormai apportato indiscussi

⁷ Cfr. F. Brugère, *L'éthique du "care"*, PUF, Paris 2021. Secondo Brugère il discorso sulla “cura” si è rivelato oggetto di un processo di marginalizzazione e occultamento che, soprattutto a partire dalla modernità, lo ha estromesso dai dispositivi istituzionali dominanti. La logica di stampo liberale con cui si è andata configurando la nostra società dall’età moderna alla contemporaneità, ha legittimato la diffusione di una “grande narrazione” liberale per disporre l’ordine collettivo sulla base del dominio di un soggetto autonomo, libero, indipendente, razionale e prestazionale. Viceversa, i tratti costitutivi della vulnerabilità e della interdipendenza degli individui sono stati immancabilmente marginalizzati.

⁸ Riprendendo la classificazione dei robot utilizzati nell’ambito della cura per le persone anziane di Amanda Sharkey, questa si suddivide in tre categorie: a) *Assistive robots and robotics*; b) *Monitoring and supervising robots*; c) *Companion robots*, letteralmente “robot da compagnia”, sui quali verterà il nostro interesse. Cfr. A. Scharkey, *Robots and Human Dignity. A Consideration of the Effects of Robot Care on the Dignity of Older People*, in “Ethics and Information Technology”, 16 (1), 2014, pp. 63-75.

⁹ Cfr. M. Bevilacqua et. al., *Social Robot to Support Older People With Dementia. A Study Protocol With Paro Seal Robot in an Italian Alzheimer's Day Center*, in “Frontiers in Public Health”, 11, 2023, ultimo accesso 13/11/2024. DOI: 10.3389/fpubh.2023.1141460.

vantaggi e miglorie ai pazienti o agli assistiti, attestati anche da ricerche scientifiche ed accademiche¹⁰.

In breve, alcuni esempi: gli animali domestici *Joy for All* (cani e gatti) dell'azienda statunitense *Ageless Innovation*, pensati per fare compagnia agli anziani con Alzheimer o con una qualsivoglia demenza senile; *ElliQ*, progettato da una società israeliana per limitare e risolvere per quanto possibile il problema della solitudine e dell'isolamento sociale e che, come affermato dal CEO di *Intuitions Robotic*, “[ElliQ] dà all'utente finale la sensazione che l'empatia sia proiettata verso di lui e risponda al bisogno sociale umano di ‘esser visto’”¹¹. Il suo successo dipende dapprima dalla capacità di instaurare una “relazione” con il suo utente, calibrando sul “riconoscimento” di questo la conversazione, per poi doverlo interrompere e ricordargli che lo stesso robot non è reale, pronunciando la proposizione: “io non sono reale”. Dalla *Softbank Robotics* è nato poi *Pepper*¹², un robot umanoide progettato per convivere con gli esseri umani, capace di conversare e reagire alle emozioni o agli stati d'animo di chi gli sta vicino. Oltre agli agenti di compagnia sociale, recentemente è stata sviluppata una tecnologia di compagnia predittiva: *Robbie*, un robot umanoide progettato presso l'*Edge Hill University* di Lancaster e addestrato ad individuare segnali di depressione e ansia nelle persone affette da demenza.

Benché diversi studi abbiano dimostrato effetti positivi derivanti dall'interazione uomo-macchina sul benessere psicologico, sociale ed emotivo delle persone anziane o con disabilità¹³, venendo queste anche coinvolte in attività strutturate o complesse, numerose sono state le difficoltà in cui le indagini sperimentali sono incorse: prima fra tutte, quella percepita dagli utenti nell'interagire direttamente con il robot senza supporto dell'operatore umano¹⁴. Più precisamente, le critiche, rivolte ai

¹⁰ Per una *overview* dei robot sociali sviluppati in Giappone, negli USA e nel Regno Unito, poi commercializzati anche in UE, cfr. G. Pratt, C. Johnston, K. Johnson, *Robots and Care of the Ageing Self. An Emerging Economy of Loneliness*, in “Economy and space”, vol. 55 (8), 2023.

¹¹ “[ElliQ] gives the end user the feeling that empathy is projected towards them and answers the human social need ‘to be seen’” (M. Osborne, *New York State Purchases Robot Companions for the Elderly*, in “Smithsonian Magazine”, 2022, ultimo accesso 10/11/2024, <https://www.smithsonianmag.com/smart-news/new-york-state-purchases-robot-companions-for-the-elderly-180980181/>).

¹² S. Nussey, *Exclusive SoftBank Shrinks Robotics Business, Stops Pepper Production Sources*, in “Reuters”, 28/06/2021, ultimo accesso 12/11/2024, <https://www.reuters.com/technology/exclusive-softbank-shrinks-robotics-business-stops-pepper-production-sources-2021-06-28/>).

¹³ H.S Kang, K. Makimoto, R. Konno, I.S Koh, *Review of Outcome Measures in PARO Robot Intervention Studies for Dementia Care*, in “Geriatric Nursing”, 41(3), 2020, pp. 207-214.

¹⁴ F. Carros, *et al.*, *Exploring Human-Robot Interaction with the Elderly*, in “Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems”, 2020, pp. 1-12.

robot di cura e sorte nell'ambito della bioetica e dell'etica della tecnologia, tendono a concentrarsi su tre problematicità: la probabilità che l'impiego di agenti artificiali in ambito assistenzialistico esacerbi un ulteriore aumento dell'isolamento sociale nelle comunità di anziani e/o disabili; dubbi sulla capacità delle macchine di fornire ai loro utenti delle cure moralmente buone; la possibilità che il loro impiego causi una netta diminuzione dell'assistenza umana¹⁵.

Potendo distinguere alcune specifiche funzioni per le quali i robot vengono addestrati, come il perseguire *attività volte ad uno scopo* oppure *orientate alla pratica*¹⁶ (distinzione non sempre così netta nella realtà), l'approccio teorico delle capacità di Nussbaum e Sen potrebbe, in un primo momento, giudicare quanto l'utilizzo della tecnologia possa promuovere il benessere dell'interessato. Se quest'ultimo dipendesse soltanto dalle possibilità che gli individui avrebbero nel poter esercitare alcune "capacità fondamentali" (come il muoversi liberamente, il comunicare, l'intrattenere relazioni etc.), allora non vi sarebbe alcun ripensamento circa l'impiego degli *assistive robots* e dei robot sociali atti al supporto e all'assistenza di individui vulnerabili. Tuttavia, come sostenuto da Coeckelbergh¹⁷, vi sarebbero alcune incrinature. Nell'approccio delle capacità mancherebbe una valutazione del *modo* in cui l'interazione con i robot trasformi proprio l'esercizio di quelle stesse "capacità", così come i contesti e le pratiche annesse. A tal proposito, Santoni de Sio e Van Wynsberghe¹⁸ prendono in considerazione una semplice attività diretta ad uno scopo: il trasportare, sollevando, la persona anziana da una sedia a rotelle al letto, dunque da un posto all'altro. Immaginando, in tale azione, la sostituzione dell'essere umano con una macchina, ecco che il risultato sarà sempre lo stesso, non vi saranno perciò evidenti alterazioni. Tuttavia, questa attività, così tanto semplice ed elementare, può

¹⁵ Cfr. A. Sharkey, N. Sharkey, *The Eldercare Factory*, in "Gerontology", 58 (3), 2012, pp. 382-388; R. Sparrow, L. Sparrow, *In the Hands of Machines? The future of Aged Care*, in "Mind and Machine", 16, 2006, pp. 141-161; M. Coeckelbergh, *Artificial Agents, Good Care, and Modernity*, in "Theoretical Medicine and Bioethics", 36 (4), 2015, pp. 265-77. Interessante è anche la proposta di Mercer E. Gary (*Care Robots, Crises of Capitalism, and the Limits of Human Caring*, in "International Journal of Feminist Approaches to Bioethics", University of Toronto Press, Toronto 2020) che colloca e motiva le tre critiche (sopra citate) nell'utilizzo dei robot sociali nella crisi odierna del capitalismo neoliberista, che ha indotto un disimpegno umano nel lavoro di cura, nonché una mercificazione di questa.

¹⁶ F. Santoni de Sio, A. Van Wynsberghe, *When Should We Use Care Robots? The Nature-of-Activities Approach*, in "Science and Engineering Ethics", 22, 2016, pp. 1745-1760.

¹⁷ M. Coeckelbergh, *How I Learn To Love The Robot: Capabilities, Information Technologies, and Elderly Care*, in I. Oosterlaken, J. Van Den Hoven (a cura di), *The Capabilities Approach, Technology and Design*, Springer, Dordrecht 2012, pp. 77-86.

¹⁸ Cfr. F. Santoni de Sio, A. Van Wynsberghe, *When Should We Use Care Robots?*, cit., pp. 1745-1760.

essere descritta nei termini di un'attività di cura che non solo è diretta ad uno scopo ma anche alla possibilità di costruire, nutrire una relazione di fiducia o un momento di socialità. La prossimità tra i due, data anche dal contatto fisico e dalla capacità del *caregiver* di divenire un "buon ascoltatore" nel saper cogliere e corrispondere alle specificità richieste dall'assistito (non autosufficiente o con bisogni assistenziali più o meno complessi), divengono tutte condizioni di possibilità di fioritura di una relazione di cura¹⁹.

Il problema fondamentale – e, secondo noi, il più urgente – consiste nella valutazione di quanto la sostituzione di un essere umano con una macchina (sia questa un robot umanoide o un pet-robot) possa inficiare e sottrarre valore a quegli elementi fondamentali tradizionalmente associati alla pratica di cura, quali capacità di attenzione, responsabilità, competenza, reciprocità, rispetto e molti altri. Secondo la distinzione operata da Sparrow tra quei servizi e/o attività di supporto di cui i robot si fanno carico, dunque i servizi fisici alla persona e il lavoro emotivo di cura, i robot non sarebbero in grado di svolgere questo secondo ordine di servizio, infatti: "interiorità e capacità di entrare in insiemi ricchi di relazioni affettive (che sono costituite dalla reciproca vulnerabilità e dalle proprietà contingenti del fatto che l'uomo è incarnato) sono necessarie per stabilire queste relazioni etiche"²⁰. Soltanto persone dotate di un corpo e della capacità di avere emozioni, nonché di interagire socialmente in una relazione duale, sarebbero garanti (e non in tutti i casi, peraltro) di valori quali cura, fiducia, rispetto e riconoscimento.

Quanto, dunque, l'introduzione degli agenti sociali artificiali sconvolge la qualità e la natura stessa della relazionalità sociale umana? Tale sconvolgimento è evitabile?

3. La *sintonizzazione* dell'Altro robotico. Una famiglia "allargata"

La costruzione di agenti artificiali atti a svolgere il ruolo di partner sociali per agenti umani sconvolge, da un lato, la natura stessa delle nostre relazioni sociali, introducendo una pluralità di agenti partecipi in una relazione e dall'altro, proprio l'introduzione di nuovi attori sociali comporta una riflessione sul *chi siamo* e sul nostro *vivere insieme*. Nonostante molti di questi strumenti tecnologici si trovino ancora in fase di

¹⁹ Non rientra nel nostro intento demonizzare qui la pratica o l'efficacia terapeutica dell'impiego dei robot sociali in ambito assistenziale, ma considerare gli orizzonti di possibilità dati dalla loro interazione con l'utente umano.

²⁰ R. Sparrow, *Robots in Aged Care: A Dystopian Future?*, in "AI & Society", 31, 2016, p. 448.

sviluppo, di perfezionamento e, in alcuni Paesi di impiego, risultano già accomunati da una forte “presenza sociale” e dalla capacità di interagire con gli interlocutori umani, rivolgendo loro *attenzioni*. Tuttavia, si potrebbe effettivamente parlare di un miglioramento della condizione degli anziani proponendo loro un’interazione con i robot, oppure, in alcuni casi – come nei pazienti con una forma avanzata di demenza – lasciando loro credere di avere davanti a sé non delle macchine, ma delle persone o animali in carne ed ossa? Si può in tal caso parlare di *inganno* e/o di uno scambio emozionale con il proprio utente inautentico?

Data l’importanza della pratica “incarnata” per la comprensione dell’assistenza, l’interazione diretta e corporale con entità “fittizie” esporrebbe l’utente ad una maggiore vulnerabilità e dipendenza dalle risposte emotive del robot nonché a danni eticamente significativi. Essendo i robot sociali delle macchine, nonché artefatti robotici o entità materiali le cui caratteristiche sono sempre il risultato di un processo di progettazione, somiglianti, in taluni casi, all’essere umano e capaci di *simulare* le apparenze o i comportamenti di questo, essi possono generare nell’utente l’impressione di interagire con partner sociali “reali”, questione certamente non immune da criticità²¹. D’altra parte, sottratto il primato tradizionalmente accordato all’esperienza soggettiva e privata del conoscere e, dunque, alla tesi tradizionale dell’omogeneità del dominio cognitivo umano, un ramo dell’attuale robotica sociale sta perseguendo un’originale linea esplorativa che si concentra sulla dimensione sociale, esterna e pubblica delle emozioni²². La questione, oggi, riguarda l’opposizione tra emozioni umane ritenute *genuine* ed emozioni robotiche *simulate ed inautentiche*; opposizione da cui trae origine la diffidenza dal coinvolgimento di esseri umani in interazioni emozionali con agenti robotici, ancor di più se utilizzati a sostegno di individui vulnerabili, come – nel nostro caso – degli anziani. L’urgenza di risolvere tale dicotomia ha portato la robotica sociale ad un avanzamento, consistente nell’implementazione dell’interazione tra “aspetti interni” e “aspetti esterni”²³ delle emozioni,

²¹ Cfr. M. Coeckelbergh, *Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability*, in “Science and Engineering Ethics”, 2019, pp. 4-5; J. Borenstein, Y. Pearson, *Robot Caregivers: Harbingers of Expanded Freedom for All?*, in “Ethics and Information Technology”, 12 (3), 2010, p. 283.

²² L. Damiano, P. Dumouchel, H. Lehmann, *Towards Human-Robot Affective Co-evolution Overcoming Oppositions in Constructing Emotions and Empathy*, in “International Journal of Social Robotics”, 7 (1), 2015, pp. 7-18.

²³ Gli aspetti “esterni” sono associati alla dimensione sociale dell’emotività, dunque all’espressività emozionale esterna, quelli “interni” sono associati alla dimensione individuale dell’emotività, nonché alla regolazione emozionale interna. La loro rigida differenziazione teorica (Cfr. M.A. Arbib, J.M. Fellous, *Emotions. From Brain to Robots*, in “Trends in Cognitive Science”, 8 (12), 2004, pp. 554-561), prevalente nella storia della filosofia occidentale e poi ereditata dalla scienza cognitiva classica, legittima nella ricerca robotica

nonché ad un confronto con lo studio neuroscientifico di queste e con gli sviluppi teorici dell'approccio radicalmente incarnato (*embodied*) alla cognizione²⁴. Dall'ipotesi scaturita emerge che le emozioni sono prodotti di un processo continuo di coordinazione tra uno spazio *intra* ed *extra* individuale, perciò risultanti da un'interazione dinamica tra questi e non esclusivamente fenomeni privati o prettamente interni. In questo modo, la questione che si pone non è più quella di scegliere o creare robot con emozioni "reali" corrispondenti a stati interni, simili al meccanismo emotivo umano, ma quella di stabilire se e quanto i robot (ed è questa la tesi della "coordinazione affettiva"²⁵) siano in grado di partecipare ad una dinamica interattiva in grado di modellare, sulla base dell'espressione affettiva, l'azione degli agenti artificiali in relazione a quella dei loro fruitori umani e viceversa.

Un simile avanzamento nella ricerca consente una riflessione etica capace di non condannare a priori tutta la produzione robotica e di poter fare spazio ad una nuova possibilità, ossia la creazione di una "zona grigia" in cui i robot si farebbero carico di attività di cura nonostante non possiedano una vita mentale alla pari di quella riconosciuta agli esseri umani, così come una analoga capacità emotiva. Infatti, l'ipotesi della coordinazione affettiva conferisce validità alla dimensione interindividuale delle emozioni ed al configurarsi di queste in un processo intersoggettivo basato – in tal caso – sul concetto di "reazione", ovvero "reagire in risposta"²⁶. L'esplorazione di questi territori "di mezzo" consente l'apertura di nuovi spiragli di potenzialità e di virtuosità derivanti dall'introduzione dei robot in ambito assistenziale, nonché la

l'esistenza di due approcci significativamente diversi alle emozioni e all'empatia artificiali. L'area di indagine riferita agli aspetti *esterni o sociali* delle emozioni ammette la costruzione di robot che sfruttano l'inclinazione umana all'antropomorfismo al fine di allacciare relazioni emozionali con gli utenti; l'area di investigazione associata agli *aspetti interni o individuali* delle emozioni è dedicata alla costruzione di agenti robotici dotati di apparati di regolazione emozionale di ispirazione biologica. Per un approfondimento inerente la dicotomia "interno"/"esterno", poi affievolita dalla più recente ricerca robotica, vedi L. Damiano, P. Dumouchel, H. Lehmann, *Dicotomie instabili. Emozioni ed empatia sintetiche*, in "Riflessioni sistemiche", 8, 2013, pp. 5-18.

²⁴ L. Damiano, P. Dumouchel, H. Lehmann, *Dicotomie instabili. Emozioni ed empatia sintetiche*, cit., p. 9. L'emergere dell'approccio interazionista del circuito affettivo uomo-macchina ha trovato sostegno tanto nelle scienze cognitive, nello specifico all'interno della scuola dell'incorporazione radicale, tanto nell'indagine neuroscientifico, impegnata nello studio dei *meccanismi di mirroring* (ovvero ai "neuroni specchio", individuati nella regione F5 della corteccia premotoria del macaco, e all'analogo sistema individuato nell'uomo). Cfr. G. Rizzolatti, L. Fogassi, V. Gallese, *Neurophysiological Mechanisms Underlying the Understanding and Imitation of Action*, in "Nature Reviews Neuroscience", 2, 2001, pp. 661-670.

²⁵ L. Damiano, P. Dumouchel, *Vivere con i robot*, cit., p. 28.

²⁶ Ivi, p. 136.

plausibilità di una diversa forma di cura valutata non tanto sulla base delle capacità di chi la fornisce quanto sulle aspettative di chi la riceve e dalla peculiare relazione che prenderà forma. Un rapporto di fiducia, ad esempio, presuppone la presenza di almeno due esseri umani, dove ciascuno di questi si aspetterà dei comportamenti o delle intenzioni non-malevoli nei propri confronti da parte dell'altro. Questo tipo di fiducia poggia su genuine emozioni insite negli individui e sul loro reciproco riconoscimento, motivo per cui sarebbe – ad oggi – impossibile rilevare questo tipo di relazione tra un agente artificiale e un agente umano. Tuttavia, in base alle ricerche realizzate nel settore della robotica – attualmente impegnata nella progettazione di robot capaci di creare dei “loop affettivi” e di coinvolgere in misura sempre maggiore l'essere umano grazie alle proprie espressioni e *reazioni* – non è da escludere che una simile relazione di fiducia possa instaurarsi con una macchina. Certamente, ci troveremo dinanzi ad una relazione di fiducia con caratteristiche diverse da quella interpersonale²⁷, data la peculiarità degli agenti coinvolti, ma si tratterà ugualmente di una relazione capace di rispondere positivamente alle aspettative o alle esigenze del singolo. Nonostante la diversità dalla più tradizionale relazione umano-umano, anche quest'altra interazione può essere moralmente rilevante ed emotivamente ricca, piena di senso per chi la esperisce.

Un'ultima considerazione. Analizzando, qui, alcune delle più recenti strade percorse dalla ricerca nel campo della robotica, non intendiamo asserire come quelle criticità etico-morali derivanti dall'integrazione uomo-robot nell'ambito della cura e dell'assistenza agli anziani siano state portate a compimento. Intendiamo piuttosto avviare una riflessione sulla possibilità di dover ripensare le caratteristiche attribuite alla relazione di cura, essendo necessaria l'ammissione di un allargamento delle nostre pratiche morali, data la pluralità e l'eterogeneità degli agenti coinvolti.

²⁷ J. Kirkpatrick, E.N Hahn, A.J Hoffler, *Trust and Human-Robot Interactions*, in P. Lin, R. Jenkins, K. Abney (a cura di), *Robot-Ethics 2.0*, Oxford University Press, Oxford 2017, pp. 146-150. La relazione di fiducia tra sistema tecnologico e utente umano risente inevitabilmente del problema dell'*inganno robotico* e dunque della legittimità morale di una simile relazione. Problema, da cui si origina da un lato il principio secondo cui gli inganni robotici possano essere giustificati in base ai benefici che comportano all'ingannato; dall'altro, l'obiezione a tale principio per cui l'inganno potrebbe essere definito come non moralmente accettabile a prescindere dagli effetti apportati. In realtà, a risultare problematiche sono ambedue le prospettive poiché intendono l'utente umano in un senso prettamente passivo, ossia come vittima ignara dell'inganno perpetrato dall'agente robotico. A tal proposito, cfr. R.C. Arkin, *Ethics of Robotic Deceptio*, in “IEEE Technology and Society Magazine”, 37 (3), 2018, pp. 18-19, DOI: 10.1109/MTS.2018.2857638.

4. L'importanza del non-visibile

Nella cornice qui descritta, i robot sono ritratti come “sostituti” degli esseri umani la cui funzione è quella di supportare quotidianamente persone malate, anziane o disabili, permettendo loro di guadagnare quell'autonomia che hanno visto progressivamente ridursi oppure limitarsi così come di ottenere valide alternative all'isolamento sociale e alla solitudine. Al fine di formulare dei piani di assistenza individuali, nell'obiettivo di cogliere le specificità e i bisogni del singolo, è stato ritenuto idoneo l'utilizzo della scala di Barthel²⁸, atta a valutare il grado di autonomia della persona anziana nelle attività quotidiane. In questo senso, l'utilizzo dei social robot nelle RSA, in contesti ospedalieri o in una relazione diadica, a “tu-per-tu” con il fruitore (dunque ad utilizzo “personale”²⁹), potrebbe effettivamente rappresentare un valido aiuto in difesa di quel grado residuo di autonomia dapprima valutato dagli specialisti mediante appositi strumenti di controllo. Oltre a fornire compagnia agli anziani (e a svolgere altre specifiche funzioni che abbiamo già menzionato nel precedente paragrafo) come rimedio alla solitudine, i robot affiancano l'utente – il cui status di persona è precario, compromesso o incerto – al fine di allinearli, per quanto possibile, all'immagine di un soggetto autonomo e indipendente, capace di “fare da sé”. Qual è, allora, il posto occupato dalla cura in una società che premia e propaga l'autonomia, il successo, la produttività, il benessere individuale, etc. e che vede nei loro opposti i segni del fallimento, della vulnerabilità, del pericolo?

Questa narrazione, sebbene “potente”, poiché con radici profonde nel pensiero moderno³⁰, ci ha consegnato un senso di socialità distorto. Il fine o il motivo dell'essere in società altro non è se non la preservazione del benessere del singolo, piuttosto che la fioritura di ciascuno all'interno di un tessuto relazionale.

Ciò che qui proponiamo è un ripensamento della convivenza umana, che ci costringe ad un confronto necessario – individuale e collettivo – con la malattia, con la nostra fragilità e vulnerabilità, condizioni comuni

²⁸ F.I. Mahoney, D.W. Barthel, *Functional Evaluation. The Barthel Index*, in “Maryland State Medical Journal”, 14, 1965, pp. 61-5. Per un suo recente aggiornamento, cfr. A. Sainsbury *et al.*, *Reliability of the Barthel Index When Used With Older People*, in “Age Ageing”, 34 (3), 2005, pp. 228-232.

²⁹ Si veda, per esempio, il progetto *Feelix Growing*, finanziato dall'UE, in “CORDIS. Risultati della ricerca dell'UE”, 18/04/2008, ultimo accesso 13/11/2024, <https://cordis.europa.eu/article/id/29353-eu-project-to-bridge-emotional-gap-between-robots-and-humans/it>

³⁰ Tradizione che ha ritenuto l'essere tutti ugualmente aggressivi e feribili dagli altri (basti pensare a Hobbes, oppure alla riflessione liberale successiva, che trova la sua origine in Locke) la condizione da cui dover prendere distanza mediante il patto sociale e l'istituzione di un governo, frutto di una libera scelta di individui indipendenti.

a tutti gli esseri viventi, a cui nessuno può sottrarsi. La nostra vita e il nostro benessere dipendono, in positivo o in negativo, dalla presenza degli altri, siano questi vicini oppure lontani e, potremmo oggi aggiungere, agenti umani oppure agenti artificiali. Se la pandemia di Covid-19 è stata un'occasione per rimettere a tema l'esperienza complessa e contraddittoria dell'interdipendenza e il tema del rapporto con l'altro, la nostra proposta è estendere tale riflessione morale alla relazione uomo-macchina in ambito assistenzialistico. Come considerare l'Altro-robotico? Una risorsa o una minaccia per l'autonomia, l'indipendenza ed il benessere degli anziani, posti ai margini da una società sempre "di corsa"?

Una cosa è certa: l'integrazione tra le cosiddette ecologie sociali della dimensione robotica con l'utente umano, ancor più se quest'ultimo è ritenuto particolarmente vulnerabile (come nel caso dei soggetti anziani), conduce necessariamente ad un'indagine volta ad approfondire una nuova forma di relazionalità non più occultabile né residuale, tanto da trovarsi oggi tra le innumerevoli sfide del futuro. Da un lato, dunque, un organismo dotato di un corpo che dispone di intenzioni, paure, attese, volontà, desideri ed inserito in un ciclo vitale; dall'altro, un meccanismo che, sfruttando illusioni zoomorfiche (ed è quanto avviene nel caso dei robot sociali dotati di caratteri corporei, linguaggio parlato, riconoscimento ed espressione delle emozioni etc.), persegue lo scopo per cui è stato fabbricato. Tra i due (utenti e robot) si innesta una relazione di fiducia, in cui l'"inganno" – scaturito dall'artefatto tecnologico – non diviene un ingrediente aggiunto o sottratto a piacimento, bensì una sua "componente strutturale"³¹ avente, nella maggioranza dei casi, ripercussioni del tutto positive sul benessere umano³². Nel delicato contesto della cura, l'approccio di carattere consequenzialista che, rispetto a posizioni "puritane"³³ reagisce meglio alle sollecitazioni poste dalla robotica, si schiera a favore di un utilizzo virtuoso dell'inganno nella dinamica umano-agente artificiale. Ne è un esempio quanto sostenuto da Andreas Matthias in *Robot Lies in Health Care*, per il quale piccole bugie a fin di bene rappresenterebbero uno strumento relazionale comune, venendo i

³¹ S. Natale, *Macchine ingannevoli. Comunicazione, tecnologia, intelligenza artificiale*, Einaudi, Torino 2022, p. 71.

³² Oggi, come già accennato nel presente articolo, vi sono numerosi esempi di progetti di robotica sociale che intendono impiegare i nuovi robot come mediatori sociali e/o terapeutici per persone con uno stato di fragilità particolarmente accentuato. Cfr. X. Wang, J. Shen, Q. Chen, *How PARO Can Help Older People in Elderly Care Facilities: A Systematic Review of RCT*, in "International journal of nursing knowledge", 33 (1), 2022, pp. 29-39; T. Tanioka, *Nursing and Rehabilitative Care of the Elderly Using Humanoid Robots*, in "The Journal of Medical Investigation", 66 (1.2), 2019, pp. 19-23. DOI: <https://doi.org/10.2152/jmi.66.19>.

³³ Cfr. W. Wallach, C. Allen, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, Oxford 2008, p. 44.

robot ad abitare un contesto di cura, dunque, ad “ingannare” per il bene dei pazienti³⁴.

Rifutando la netta contrapposizione tra tecnofilia e tecnofobia, è altresì vero che la creazione di un robot che “non serve a niente di particolare e può fare qualsiasi cosa”³⁵ diviene, al contempo, uno strumento utile per comprendere meglio *chi* siamo e come interagiamo tra di noi, nonché lo spazio di azione che il nostro reciproco dipendere gli uni dagli altri ci consegna. Spazio che comporta inevitabilmente una forma di responsabilità, un impegno epistemologico consapevole tale da consentire ai bisogni di chi è intorno a noi di affiorare e di farsi visibili, di divenire dicibili e vivibili considerando la diversità delle nostre vite.

Senza voler disconoscere la valenza terapeutica dei robot sociali così come l'avanzamento della progettazione di questi da parte dell'attuale ricerca robotica, non va dimenticato che, oltre ad essere *presenti* socialmente (come agenti fisici ed esistenti a tre dimensioni), questi robot sono in grado di rivolgere la propria *attenzione*³⁶ agli altri, di avere cura “per il particolare” e – nel nostro caso – di riconoscere quella parte della nostra società (di certo non l'unica) che abbiamo destinato all'invisibilità e al silenzio, gli anziani. Volendo lasciare da parte la *vexata quaestio* della filosofia morale e politica relativa alle condizioni in cui sia eticamente giustificabile ingannare altri esseri umani – laddove il veicolo dell'inganno è l'oggetto robotico e l'ingannato l'utente – potrebbero i robot sociali, per come volutamente progettati, sopperire ed implementare un'incapacità umana, quale l'attenzione per ciò che Veena Das chiama *la vita quotidiana dell'essere umano*³⁷, ossia, l'attenzione verso fenomeni solitamente non visti, ma che sono costantemente sotto i nostri occhi?

Lungi dalla pretesa di fornire risposte risolutive in merito, vorremmo dapprima ri-conoscere l'importanza di un significato fondante l'etica della cura³⁸ (oltre alla già citata attenzione per il quotidiano), e cioè che *la per-*

³⁴ A. Matthias, *Robot Lies in Health Care: When Is Deception Morally Permissible?*, in “Kennedy Institute of Ethics Journal”, 25 (2), 2015, p. 179.

³⁵ L. Damiano, P. Dumouchel, *Vivere con i robot*, cit., p. 43.

³⁶ Ivi, p. 22.

³⁷ V. Das, S. Cavell, *Life and Words: Violence and the Descent into the Ordinary*, University of California Press, Berkeley 2007.

³⁸ Il riferimento implicito è rivolto principalmente a quel filone riflessivo sulla cura che trova come riferimento lo studio pionieristico di C. Gilligan, *In a Different Voice. Psychological Theory and Women's Development*, Cambridge, Harvard University Press 1982, dal quale sono dipartite differenti linee di sviluppo (consonanti e discordanti) che considerano la cura come nozione centrale di un nuovo paradigma morale da contrapporre al paradigma dominante di un'“etica della giustizia”. Dunque, una linea riflessiva incentrata su una tematizzazione della morale che ruota intorno alla capacità di prendersi cura delle reti di relazioni in cui ci troviamo immersi (e che in quanto tali potrebbero non essere frutto di una nostra scelta). Tra le autrici e gli autori che hanno ereditato e ulteriormente

sona è vulnerabile – tratto caratterizzante di una condizione che ci accomuna; in seguito, ricordare quale sia l’obiettivo di questa: riportare l’etica verso l’ordinario³⁹, nella promessa di non considerare nessuno trascurabile. La cura diretta all’importanza del non visibile, la consapevolezza di un senso morale che non è mai fisso ma sempre particolare, l’allargamento dei concetti morali che solitamente applichiamo alla realtà divengono il perno di una differente visione che consiste nel cogliere le possibilità o i significati che emergono dalle cose, dalle situazioni o dai momenti.

A tal proposito, il nostro articolo vorrebbe avere una duplice finalità. Essendo il senso dell’esperienza ordinaria mai *dato* a priori – sostiene Cora Diamond⁴⁰ – ma costantemente ridefinito dalle nostre pratiche morali “esplorative”, queste dovrebbero estendersi oltre le tradizionali interazioni umano-umano, per riconoscere nuove e/o diverse possibilità relazionali (il riferimento è sempre alla relazione di cura e di assistenza uomo-macchina), altrettanto ricche e moralmente rilevanti. In questo senso, è ingenuo e rischioso ritenere la capacità di cura una prerogativa esclusivamente femminile così come esclusivamente umana (basti pensare agli animali non umani e, in tal caso, agli artefatti artificiali appositamente progettati).

Condizione di possibilità dell’allargamento del tessuto relazionale umano, che coinvolge ora la partecipazione di nuovi agenti sociali, è un confronto con l’incertezza, con un “disequilibrio percettivo” capace di portarci oltre il “rifiuto dell’avventura morale”⁴¹ che spesso giustificiamo per la nostra difficoltà sostanziale nel voler conoscere gli “altri” – compreso l’Altro robotico – o nel mantenere un atteggiamento prettamente scettico dinanzi la novità. Ciò non significa sottovalutare i pericoli derivanti dall’introduzione dei robot sociali nel nostro spazio abituale e relazionale, potendo questi de-responsabilizzare l’umano nelle attività di assistenza e di cura ma considerarci “inesperti” nella vita morale per accogliere le “parole” di qualcuno che sembrano essere estranee al nostro mondo concettuale condiviso.

Attestata la cura come cura delle relazioni e pratica che si dà nelle relazioni e stabilito il nostro essere contagiabili e bisognosi di atten-

sviluppato (chi nei tratti pedagogici, chi nei tratti etici o ancora in quelli più spiccatamente politici) la prospettiva di C. Gilligan, vi sono: N. Noddings, J.C. Tronto, F. Kittay, ma anche A. Baier e M. Slote. In area italiana, sono da segnalare i lavori di E. Pulcini, con le sue riflessioni sulla cura calate nello scenario globalizzato attuale, e quelli di L. Mortari, sensibile a un approccio di stampo ermeneutico-fenomenologico alla cura.

³⁹ Cfr. J.C. Tronto, *Préface à l’édition française*, in Id., *Un Monde vulnérable. Pour une politique du care*, La Découverte, Paris 2009.

⁴⁰ C. Diamond, *The Realistic Spirit. Wittgenstein, Philosophy and Mind*, The MIT Press, Cambridge 1991, p. 27.

⁴¹ Ivi, pp. 314-315.

zioni, entrambi segni della nostra vulnerabilità, ecco che l'irruzione della cura nello spazio pubblico, oggi attribuita, in parte, ai cosiddetti "sostituti", dovrebbe non ritenere l'autonomia come elemento da preservare ed incrementare grazie all'impiego dei robot sociali – a loro volta programmati per portare i loro utenti ad aderire a norme e principi universali, imparziali e razionali. La cura dovrebbe invece ripartire dalla capacità di riconoscere la concretezza, la particolarità e la diversità di ciascuno, per poi prendersene carico, così come dall'attenzione ai bisogni, alle vulnerabilità degli anziani, nei cui confronti vi è spesso una disattenzione morale. Disattenzione che, pensiamo, potrebbe essere parzialmente risanata attraverso l'invito ad esercitare e coltivare la nostra capacità di riconoscere le forme innumerevoli dei legami, delle connessioni e delle relazioni che abitano il nostro universo morale, per ri-descrivere il mondo in cui viviamo e riconoscere la potenza della realtà umana. Se è vero, come ci insegna Hannah Arendt, che quest'ultima sia da individuare nella sua radicale *pluralità*⁴² – presupposto e condizione che le permette la descrizione di una soggettività dialogica, che esige la reciprocità e la comunicazione per dirsi e dotarsi di significato – ne consegue che l'integrazione di nuovi artefatti sociali nel tessuto relazionale umano non potrebbe che arricchire questa potenza.

Allora, potrà mai divenire il ricorso agli agenti robotici sociali nella relazione di cura – ed questa la seconda finalità del presente articolo, nonché un sollecito diretto a noi stessi – un ausilio nel trasformare la voce privata, flebile, da noi inascoltata delle persone anziane, per renderla così espressiva da destare finalmente la *nostra* attenzione nei loro confronti?

⁴² Per comprendere la premessa a fondamento della sua analisi antropologica, che riconosce la pluralità come condizione ontologica dell'essere umano, cfr. H. Arendt, *La vita della mente*, il Mulino, Bologna 2009, pp. 99-102.

*Simona Tiribelli**

Artificial Intelligence and Moral Progress: A Missing Debate

Abstract

This article addresses an underexplored topic in the field of ethics of artificial intelligence (AI), that is, the impact of AI technology on moral progress at the collective level, and shows that such an inquiry is not only ethically relevant but necessary for the design of AI systems able to protect our moral gains and sustain the possibility of further societal moral progress in our contemporary information societies increasingly shaped if not yet governed by AI technology.

Keywords

Ethics of Artificial Intelligence (AI); Societal Moral Progress; Collective Moral Reasoning; Epistemic Agency; Ethics by Design

1. Moral progress and AI ethics

Over the last decade, a great deal of philosophical work has been focused on identifying, assessing, and addressing the ethical risks and the controversial societal implications of the fast-paced design, deployment, and use of AI technology. The main approach underlying such work, which is structured today in the specific applied ethics field known as AI ethics, is a *proactive one*: it is grounded on the idea that we should develop specific ethical knowledge or apply ethical theories as developed in moral philosophy and applied ethics to prevent AI ethical risks, instead of just mitigating by chasing them afterward, through an approach of *ethics by design*. Prominent research in AI ethics does not only focus on how to design AI systems that prevent further social discrimination and new human rights infringement phenomena. They focus also on how to

* Università degli Studi di Macerata.



proactively ethically design such systems to address socially rooted problems, such as unfair inequalities, and create fairer and *morally better* societies (O'Neil 2016; Benjamin 2018; Giovanola & Tiribelli 2022). Indeed, the goal of a prominent deal of scholarship in AI ethics does not lie in the design of neutral or objective AI technology, which would preserve the *status quo*. This approach would be indeed sterile and inadequate in non-ideal societies as the real ones to prevent morally undesirable outcomes; that is, for example, to avoid or mitigate that AI can reflect, perpetuate, or exacerbate sources of harm such as systemic ethnic bias, structural discrimination, and unfair social inequalities that are historically rooted and culturally embedded into our societies. Rather, it focuses on embedding and translating specific ethical values into AI design, often as *ethics by design* criteria, to develop AI systems that can fix or mitigate such sources of harm and enable only those specific technical progress and advancements that both respect and promote the moral values and social goals we care collectively. That is, only that technical progress that can preserve our moral conquests (e.g., widespread acknowledgment of fundamental rights) and promote further societal and moral progress (e.g., fairer societies through further inclusion, made possible by AI systems, or through the detection and removal of historically-rooted unfair cultural biases, thanks to AI big data processing and analytics techniques)¹.

In sum, it sounds we could rightly claim that a great corpus of AI ethics scholarship assumes – more or less implicitly – moral progress in its background or tends to it as its purpose. Nevertheless, the specific topic of moral progress is surprisingly poorly explored explicitly in this domain. Thus far, a systematic ethical inquiry based on insights from theories on moral progress to investigate whether AI systems promote or hinder the necessary preconditions for moral progress at the collective level has neither been carried out in the AI ethics debate nor in the moral progress one².

This article addresses such a gap and shows that such an inquiry is not only ethically relevant but necessary for the design of AI systems in a way to protect our moral gains and avoid phenomena of moral regress in our contemporary mature information societies that are increasingly and deeply shaped – if not yet governed – by algorithmic technology.

To this aim, the paper first clarifies the philosophical account of moral progress we adopt in our ethical inquiry, which is one of today's bench-

¹ See Tiribelli (2023) for a proposal on the use of AI capabilities to detect and mitigate sources and drivers of systemic unfair social inequalities into our societies.

² Indeed, just a few (precious) attempts have been made with regard to AI systems and moral progress that however focus on moral change at the individual level; see the analyses of Savulescu & Giubilini (2018) and Savulescu & Maslen (2015).

mark theories on moral progress, Buchanan and Powell's biocultural account of moral progress, and specifically the necessary preconditions for moral progress such account points out. Then, the paper shows how the AI systems hyperconnected into our information societies are impacting such preconditions, zooming in on their impact on the widespread exercise of open-ended moral reasoning, sketching out a few implications such impact raises. Finally, the paper concludes by calling for specific ethics by design criteria to shape the development of such AI systems so as to protect and promote the necessary preconditions securing at least at a minimum threshold the possibility for further moral progress and the basis for avoiding moral regress.

2. On moral progress

The concept of moral progress is a complex one and it is widely debated within a plurality of domains, from moral philosophy, through evolutionary psychology, to biology and cognitive (neuro)sciences. It is not within the scope of the present article to debate this concept or provide a related systematic discussion as it would lead us astray. For the purpose of the paper, however, we aim here to briefly clarify the (i) conception of moral progress we endorse in our ethical analysis considering the general debate on moral progress, (ii) the diverse types of moral progress in this sense we can contemplate, the (iii) specific account of moral progress we encompass for our inquiry, and (iv) the particular *sine qua non* preconditions for its possibility such account contends.

As it has been pointed out by Sauer et al. (2021), it is possible to distinguish at least *two* major approaches within the contemporary debate that mainly emerge on moral progress. One approach is of those who embrace a *broad conception* of moral progress, according to which moral progress can be understood as “any kind of morally desirable change” (Sauer et al. 2021, p. 2; Eriksen, 2020, pp. 5-11; Kitcher, 2017, p. 64; Sauer, 2019) from individual changes in beliefs and behaviors (e.g., a mother who learn how to be more patient with their children) to progress on a global scale (a global decline in homicide rates) to improvements in moral theory. Another approach is, instead, of those who endorse a *narrow conception* of moral progress: these scholars content that not all morally desirable changes should be considered as moral progress, but only those that are determined, for example, by the exercise or improvements of human moral capacities (Buchanan & Powell 2018, p. 51), such as moral reasoning, moral motivation, or the ability to follow moral norms (think of those moral gains in human welfare that depend on the large-scale spread of ideas about fundamental rights within some societies). In

this article, we endorse a *narrow* conception of moral progress, according to which we can consider instances of moral progress only those morally loaded phenomena that substantially depend on – or largely involve – the exercise of human *moral capacities* by a large number of people within a society. In this sense, as anticipated in the introduction of this paper, we are interested here in moral progress at the collective level, that is, moral changes that concern a widespread number of people (Buchanan & Powell 2018; Singer 1981); a number that is sufficient to elicit socio-political revolutions as those that have enabled in human history those moral conquests we refer today also as benchmark or historical examples of moral progress (e.g., the abolition of slavery)³. We specifically embrace this second view insofar as it entails the *active involvement* of the human *moral dimension* at the collective level and here we are specifically interested in understanding how AI systems can impact this human *moral dimension* and related implications for moral progress. Put it differently, we are not interested on investigating if the surge in application of AI systems may accidentally reduce some morally detrimental phenomena for reasons that have nothing to do with the exercise of practical agency of individuals and groups. Rather, we are specifically interested in investigating whether AI systems currently in use are designed and hence “operate” in such a way to promote or instead undermine the exercise of human moral capacities, such as moral reasoning, moral motivation, and the normative agency and, by doing so, affect moral progress.

In this (narrow) sense, classic examples of moral progress are historically traced, with broad agreement in the debate, to the abolition of slavery, the establishment, at least in some parts of the world, of universal suffrage, and the advancement of LGBTQ+ and animal rights. It is possible to distinguish more specific types of moral progress in its narrow conception, that is, as dependent on the exercise of human moral capacities. These types may consist of adequate *de-moralizations* that concern wrongly moralized practices in the past (e.g., sex outside marriage), or instead in *correct moralizations* of behaviors erroneously thought to be morally neutral (e.g., gender-based wage disparities); or in cases of *improvements in moral motivation* (when this comes to play a determining role on people’s behavior, leading to greater levels of compliance with moral norms) as well as in the *improvement of existing moral concepts* (for instance, the concept of responsibility thanks to that of intentionality) and in the *development of new moral concepts* (e.g., the concept of

³ We do not discuss here the nature of moral progress (if it is global, that is, only happens at the level of institutions, etc., or applies to the level of individuals) or mean it in global terms (i.e., for all humans and societies worldwide); we always refer to it here in a *localized* way (i.e., as it happens in certain socio-political contexts and societies).

sexual harassment), which enable people to recognize, understand, and communicate in an intelligible way to others a certain kind of injustice they are subject to (Fricker 2007)⁴.

Most of these types of moral progress are discussed against the backdrop of one of the most endorsed theories of moral progress in the contemporary debates, Buchanan and Powell's biocultural theory of moral progress, which is situated within the classical view of moral progress as inclusion, or gradual expansion of the circle of moral consideration (Buchanan & Powell 2018; Singer 1981). This theory is of particular interest here and argues for the possibility of moral progress and regression as dependent on the dynamic interaction between particular social environmental conditions (and the *stimuli* they generate) and the flexible or ductile (moral) cognitive capacities of individuals, which are adaptively plastic. In short: based on the stimuli that different social environments can provide, people can develop inclusive or, instead, divisive (or tribalistic) moral responses (especially morally regressive outgroup hostility) and so promote further moral progress or instead trigger phenomena of moral regress, as moral progress and conquests are not irreversible.

In particular, according to Buchanan and Powell's account, the presence of favorable socio-epistemic and institutional conditions would enable the exercise of a fundamental *necessary condition* in order to sustain moral progress⁵, in terms of large-scale socio-political changes, as those mentioned above: the *widespread exercise of critical, open-ended moral reasoning* (Buchanan & Powell 2018; Buchanan 2021, p. 44), whereby it is meant the ability of cognitively normal human beings have to make the particular moral rules and concepts they are following objects of critical scrutiny, which can sometimes lead to their modification or abandonment (Buchanan 2021, pp. 142-143). In summary, it is the "capacity to examine our moral concepts and the concept of morality itself" (Buchanan 2021, pp. 144-145)⁶. By contending how the plasticity of human cognitive capacities under favorable social (ecological) conditions can foster moral reasoning and prompt inclusive moral responses at large-scale, hence sustaining moral progress, this account acknowledges both the contributions of empirical sciences and evolutionary biology, as well as those of neuroscience (findings on cognitive bias, limits, and errors), while leav-

⁴ We refer to the work of Sauer et al. (2021) for a systematic discussion on different types of moral progress.

⁵ We take these conditions as what is required at a minimum threshold for safeguarding the chance of further moral progress; the question of whether they are also sufficient is not discussed here.

⁶ For key insights on reflexivity, consider theorists such as (classically) Horkheimer (see *Traditional and Critical Theory*, 1937) or, more recently, Jaeggi (*Critique of Forms of Life*, 2018).

ing room for the critical role of *human moral normativity* (Buchanan & Powell 2018, pp. 145-155).

Furthermore, this theory has the merit of being among the few to illuminate certain *sine qua non* conditions for moral progress at the collective level, that is, in terms of large-scale socio-political changes (i.e., involving moral improvements of a sufficiently large number of people). In short, these preconditions for moral progress are identified, first, in certain favorable *socio-epistemic conditions*; we sum up briefly a few of them as follows. There should be (i) the presence of advanced technologies to disseminate ideas and connections between ideas that enable reasoning about how people ought to live; (ii) there must be considerable freedom of expression and association so that discussion of ideas about how to live among people with diverse viewpoints can be relatively open-ended and the control over information and communication technologies (ICT) must be dispersed, so as to avoid constraints on freedom of information and expression and the curtail the exercise of critical moral reasoning; (iii) a significant number of people must be exposed to the fact that other societies have different ways of doing things (this allows for the possibility of investigating whether the moral rules they follow are optimal, the extension of sympathy beyond their group, and the identification of common interests); (iv) a developed culture of justification, of providing genuine reasons, not based, therefore, on appeals to authority or tradition. As such account points out (see Buchanan 2021, p. 149), such necessary socio-epistemic conditions depend on the character of the institutions in a society too and the sort of institutional order in which large-scale moral progress is likely to come about through peaceful means will be in broadest terms a liberal and at least minimally democratic order. Taken together, these pre-conditions are claimed to be enabling a third (moral – strictly speaking) condition, which is contended as key to the achievement of moral progress: the widespread exercise of open-ended, critical moral reasoning. Put differently: they constitute the social-epistemic contexts that are deemed necessary for the large-scale exercise of our moral capacities, and therefore, for the possible occurrence of the various types of moral progress in a narrow sense previously mentioned. For the sake of clarity, we do not proceed further on moral progress: we have indeed enough elements to carry out our ethical analysis.

In this section, we have shed light on the specific account of moral progress and the *sine qua non* conditions of moral progress we refer in order to carry out our ethical inquiry, with a specific focus on the widespread exercise of moral reasoning. In the next section, drawing on these insights, we analyze how AI systems and particularly algorithmic ICT can impact such preconditions and the implications for moral progress such impact raises.

3. Moral progress and AI systems

The impact of AI technology on the social and epistemic conditions of our agency in mature information societies has been extensively documented within the scholarship on digital ethics and ethics of AI. Early scholarship in digital ethics had already outlined how digital ICTs were reshaping the environments where we live, renewing them as novel hyperconnected, phygital, onlife ecosystems we are fully and constantly immersed in, as well as the social practices through which we develop, process, and communicate knowledge about the reality, the others, and ourselves (Floridi 2014). Later, further work has refined and expanded these arguments with special regard to the pervasive use of machine learning (ML) and deep learning (DL) algorithms that nowadays govern almost every digital ICT, from search engines to social networking services (large language model-based platforms included) and Internet of things and smart devices broadly.

The idea is catching on in this debate claims the rise of an “algorithmic governance” in today’s information societies (Yeung 2018; Tiribelli 2022), according to which ML and DL algorithms are restructuring and redefining the information-based social and epistemic environments in which we prepare and make our choices as particular individuals and as belonging to social groups. They are no longer merely “information gatekeepers” but are playing the role of new kind of choice architects (Tiribelli 2022). This is possible because, due to the ubiquitous presence of algorithm-based digital ICT, everything about reality and ourselves, from our identity characteristics, features, and onlife movements to our relationships, social affiliations, most intimate attachments, sexual orientations, attitudes, preferences, beliefs, vulnerabilities, values, and shared commitments, cannot be only described in informational terms, but can be datafied and hence practically translatable (as captured, processed, incorporated) into data and information. Everything – what we produce and what feeds us – is data and information, and the latter is processed and managed today by AI algorithms such as ML and DL.

This means that the algorithms underlying our ICTs not only decide today what information to show us, and in what order – through filtering and classification techniques – but also determine how (based on which parameters and specific objectives) our onlife *informational socio-epistemic environments* are defined (i.e., the environments where we are immersed and informationally exposed to both knowledge and relationships). In other words, these systems reshape our onlife environments in terms of informational and socio-relational exposure and availability. By doing so, they influence the socio-epistemic and moral conditions of our agency in mature information societies, by choosing what and who – both

quantitatively and qualitatively – will inform and shape the development of our knowledge (i.e., what ideas, thoughts, beliefs, relationships, etc. we will meet and will challenge our thoughts and what won't as instead non-available), including moral knowledge (i.e., what kind of moral reasons, values, and everything that can become a motive for making certain choices and taking specific actions over others will be available and what it won't).

The design of such systems, however, is oriented by parameters and objectives pre-established by third parties (i.e., technology designers and deployers), and as currently conceived, may give rise to phenomena that negatively affect at least a few of the above-mentioned socio-epistemic and moral conditions necessary for moral progress and consequently the possibility for moral progress itself.

Indeed, the ML and DL algorithms at stake here are probabilistic, meaning that they learn how to achieve predetermined goals – which today mostly (sometimes still exclusively) overlap with profit maximization through efficiency criteria – by scaling huge amounts of data to discover knowledge (i.e., patterns, correlations, etc.) that is efficiently used to categorize the users into specific groups of individuals profiled with similar identity and decision-making characteristics including similarities in cognitive distortions and biases (O'Neil 2016). These groups and hence the people silently placed in them are then presented with filtered informational (socio-relational and epistemic) contexts that are personalized for them (e.g., information and relationships aligned to the group profiled major political orientation), but specifically oriented to the efficient achievement of ML and DL goals preset by tech providers, such as maximizing clicks on third-party sponsored content (Zuboff 2018).

In doing so, however, as it has been widely documented over the last decades (Pariser 2011; Sunstein 2008, 2017; Nguyen 2020), these systems elicit some specific socio-epistemic controversial phenomena, such as the formation of filter bubbles, epistemic bubbles, or informational silos (echo-chambers), situating individuals within groups of like-minded, i.e., people with similar views, values, beliefs, practices, attitudes, interests, and preferences, as well as cognitive distortions and unfair cultural biases, where diverse, outgroup voices tend to be excluded, discredited or suppressed. Such groups and contexts are thus characterized by reduced exposure to both heterogeneous information and relationships – a systemic reduction that is difficult to experience or implement at a large scale in environments that are not governed by algorithms. It follows that these AI systems undermine rather than expand the spreading of novel or diverse ideas and viewpoints from in-group perspectives, as well as of critical, unexpected, challenging relational encounters. In this sense, they can negatively affect both the socio-epistemic conditions (such as the ne-

cessity of exposure to diverse moral ideas and social practices and the development of a culture of justification) and the moral one *stricto sensu* (i.e., the widespread exercise of critical, open-ended moral reasoning) argued to be necessary for moral progress at the collective level, raising controversial implications for both individuals, as members of specific groups, and for groups themselves, as collectives.

Indeed, in groups of like-minded people, with similar ways of thinking and acting, decision-making biases included, people tend to radicalize, rather than mitigate by mutual compensation, their own cognitive errors and cultural biases, as well as their ideas, viewpoints, and attitudes, either by social pressure dynamics or because particular cognitive bias (such as confirmation bias). As a result, in-group individuals could end up developing increasingly self-enclosed identities (Parsell 2008) with increasingly rigid or impermeable physiognomy (Tiribelli 2023), making them less able to understand and respect those who have different ideas and especially outgroup perspectives, ultimately leading to increased social cleavage and division (Parsell 2008, p. 43).

Additionally, the ingroup social pressure and emotional tendency to uncritically adhere to dominant positions or to confirm what aligns with and reinforces our pre-existing beliefs or attitudes can lead to reduced reflective scrutiny, critical thinking, and moral reasoning. In other words, it can lead to a decreased demand for – and, over time, a gradual erosion of – our exercise of moral capacities, from the ability to develop genuine moral reasons to that of both demanding and providing genuine reasons (i.e., moral accountability and answerability).

In sum, these systems, by placing individuals in environments with reduced heterogeneous informational and relational exposure, preclude them by design from encounters with different social and moral ideas, beliefs, concepts, and practices, so disabling the exercise of moral reasoning. Over time, these phenomena can raise the possibility of generating a moral reasoning deskilling, that is, a weakening of our critical ability to assess critically whether what we are adopting and following from a moral standpoint (in terms of beliefs, concepts, rules, and practices) is optimal or requires revision or change – an ability that is largely understood as critical for moral progress in the related debate.

Similar implications can be extended to groups as collectives. Indeed, it is likely that groups with increasingly self-enclosed and radicalized identities, when called to engage in critical dialogues in the public spaces of our liberal-democratic societies, will be less open to mutual understanding and agreement, as well as less inclined to cooperation in inclusive terms. They may struggle to develop shared commitments and social projects (Giovanola & Sala 2021) and hence to act together towards common goals for a better society. Rather, hard group physiogno-

mies and the lack of ingroup widely deliberated and strongly and deeply felt genuine reasons to support group identities might lead groups to become intolerantly unreasonable with each other, creating *ecological conditions* for **moral regress**.

It follows that the possibility for the exercise of open-ended, critical moral reasoning is endangered not only at the individual level but also at the collective one, as is the possibility of moral progress.

4. Conclusive remarks: Moral progress by AI design?

In this article, we have shown how AI systems governing our pervasive digital ICTs can affect the necessary conditions for moral progress by reshaping our hyper-connected informational and socio-relational environments in ways that, instead of promoting, erode the exercise of moral reasoning of both groups, as collectives, and its members, as individuals, paving the way to the possibility of losing the moral gains we care most and of incurring to phenomena of moral regress. By doing so, we have disclosed not only the relevance of such an ethical inquiry in the light of the implications it outlines; we have also shown the need of bridging the scholarship on moral progress and that of AI ethics to act to ensure AI systems become a driver instead of a hindrance of moral progress.

This necessity is even more evident if we consider prominent practical solutions aimed at promoting moral progress and avoiding moral regress elaborated in the debate on moral progress within the field of moral philosophy.

Indeed, those who contend a narrow conception of moral progress tend to adopt proactive or intentional approaches to moral progress, grounded on the possibility of eliciting moral progress or guiding societies toward moral progress, through *design* actions. Many scholars of moral progress claim that the philosophical and empirical studies on moral progress can provide precious knowledge to shape our social environments in ways that promote instead of hampering moral progress (Buchanan 2021, p. 236). However, the proposals made thus far in this debate focus only on institutional design and are mainly oriented by criteria of economic or political efficiency. In this regard, for example, Buchanan (2021) has criticized the lack of practical solutions that consider the moral effects of diverse typologies of institutional design.

Based on the ethical inquiry we sketched out in this article, we claim here that there is another gap in this regard that we might fill in to create favorable ecological conditions for moral progress.

If champions of moral progress invite to reshape or design social environments by considering knowledge on what promotes or hinders cer-

tain preconditions required by moral progress, we claim here we cannot prescind to also consider that (i) AI systems reshape such preconditions and therefore (ii) can translate into forces *for* or *against* moral progress. Thus, any design actions for promoting moral progress *should* seriously consider how intentionally re-designing interconnected AI systems considering literature on moral progress so as to make them conducive instead of adverse to it.

To sum up, in this article we hope to have shed light on the need to ethically design such systems, which means, in this case, developing design criteria that consider, in a novel way, studies on moral progress. Indeed, if such technologies show, theoretically, the potential to adversely affect moral progress, the same if redesigned considering both the *sine qua non* conditions of moral progress and the moral risks and effects that the action of AI systems on such conditions can generate hold the potential to become not only instruments of protection from moral regress, but also real forces of concrete promotion of moral progress.

Bibliography

- Benjamin, R.,
2019 *Race after technology: abolitionist tools for the new Jim code*. Polity, Cambridge.
- Buchanan, A.,
2020 *Our Moral Fate: Evolution and Escape from Tribalism*, The MIT Press, Cambridge.
- Buchanan, A., Powell, R.
2018 *The Evolution of Moral Progress. A Biocultural Theory*, Oxford University Press, New York.
- Fricker, M.,
2007 *Epistemic injustice: Power and the ethics of knowing*, Oxford University Press, Oxford.
- Giovanola, B., Sala, R.,
2021 *The reasons of the unreasonable: Is political liberalism still an option?*, "Philosophy & Social Criticism", 1226-1246.
- Giovanola, B., Tiribelli, S.,
2022 *Weapons of moral construction? On the value of fairness in algorithmic decision-making*, in "Ethics and Information Technology", 24, 1-13.
- Giubilini, A., Savulescu, J.,
2018 *The Artificial Moral Advisor. The "Ideal Observer" Meets Artificial Intelligence*, in "Philosophy & Technology", 31(2), pp. 169-188.
- Horkheimer, M.,
1937 *Traditional and Critical Theory*, The Continuum Publishing Company, New York.
- Pariser, E.,
2011 *The Filter Bubble*, Penguin, London.

- Parsell, M.,
2008 *Pernicious virtual communities: Identity, polarization and the web 2.0*, in «Ethics and Information Technology».
- Jaeggi, R.,
2018 *Critique of Forms of Life*, Harvard University Press, Cambridge.
- O'Neil, C.,
2016 *Weapons of math destruction: how big data increases inequality and threatens democracy*, Crown, New York.
- Nguyen, C.T.,
2020 *Echo Chambers and Epistemic Bubbles*, in "Episteme", 17(2), 141-161.
- Sauer, H., Blunden, C., Eriksen, C., Rehren, P.,
2021 *Moral progress: Recent developments*, in "Philosophy Compass", pp. 1-10.
- Savulescu, J., Maslen, H.,
2015 *Moral Enhancement and Artificial Intelligence: Moral AI?*. In: Romportl, J., Zackova, E., Kelemen, J. (eds.) *Beyond Artificial Intelligence. Topics in Intelligent Engineering and Informatics*, vol 9. Springer, Cham.
- Singer, P.,
(1981) 2011 *The expanding circle-Ethics, evolution, and moral progress*, Princeton University Press, Princeton.
- Sunstein, C.,
2008 *Democracy and the internet*. In: van den Hoven, J., Weckert, J. (eds.) *Information technology and moral philosophy*. Cambridge University Press, New York, pp 93-110.
- Tiribelli, S.,
2022 *Moral freedom in the age of artificial intelligence*. Mimesis International, Milan-London.
- Tiribelli, S.,
2023 *Inequalities and Artificial Intelligence*, in "Filosofia Morale/Moral Philosophy", (3).
- Tiribelli, S.,
2023 *Identità personale e algoritmi. Una questione di filosofia morale*, Carocci editore, Rome.
- Yeung, K.,
2018 *Algorithmic regulation. A critical interrogation*, in "Regulation & Governance", pp. 505-523.
- Zuboff, S.,
2019, *The age of surveillance capitalism: the fight for a human future at the new frontier of power*, Public Affairs, New York.

Recensioni

Bryan Jesus Irias Alfaro

Nunzio Bombaci, *La persona, il prossimo, l'amico.*

Le figure dell'altro in Pedro Laín Entralgo,

ETS, Pisa 2023, pp. 296

Nella letteratura spagnola del *Siglo de Oro* particolarmente in *Don Quijote de la Mancha* de Miguel de Cervantes, si può osservare la figura della persona, l'amico e il prossimo, nel personaggio di Sancho Panza e questo si evince nell'ultimo dialogo tra lo scudiere e Don Quijote allorché nel suo letto di morte Don Quijote identifica Sancho come l'amico e come il prossimo che lo accompagna in tutte le esperienze della vita, anche in quella della morte.

Il Prossimo, l'amico, l'altro... con questi diversi sostantivi che non esauriscono l'essenza dell'essere umano, Nunzio Bombaci fa riferimento alla realtà della persona umana che si presenta come altro nell'esperienza relazionale dell'essere umano. Le figure in disamina sono tratte dall'opera filosofica del filosofo spagnolo Pedro Laín Entralgo, soprattutto della sua opera *Teoría y realidad el otro*.

Nell'introduzione "all'opera" il nostro autore precisa varie cose che credo opportune sottolineare: il suo obiettivo non è "attualizzare" l'opera lainiana, piuttosto, il suo compito è quello di contestualizzarla nell'epoca e nel momento storico in cui si è evoluta; il secondo obiettivo è mettere in dialogo Pedro Laín con altri autori contemporanei.

Uno dei filosofi con cui Entralgo dialoga spesso è il metafisico Xavier Zubiri; infatti, verso la fine degli anni ottanta afferma Bombaci, si confronta soprattutto riguardo al tema del corpo e dell'anima e detta discussione come questi annuncia all'inizio dell'opera è uno dei limiti dell'opera stessa, perché tale dibattito tra i filosofi spagnoli non trova luogo in questa sede.

D'altra parte Bombaci dà molta importanza alla teoria di Julián Marías riguardo alla relazione personale e all'alterità, infatti il filosofo madrilenico asserisce che l'esperienza dell'amicizia è una realtà in cui la persona "svela" la sua intimità in quanto "*quehacer*", ovvero, da farsi, il quale porta all'istallazione con l'altro nella realtà del noi. Credo importante non tralasciare questo particolare; poiché la figura dell'*amico* accade nell'esperienza dell'amicizia, e tale figura nella teoria di Pedro Laín è una delle realtà principali un cui l'essere umano si presenta come un "tu".

Per rafforzare la tesi che Entralgo propone riguardo all'altro, Bombaci apre diverse discussioni con Cartesio, Teodor Lipps, Sartre e Heidegger e lo fa riferendosi alle teorie relazionali di Gabriel Marcel, Ortega y Gasset, Julián Marías, Zubiri e al filosofo Martin Buber, più specificatamente al suo noto scritto *Io-Tu* (1977), invertendone l'impostazione e facendola diventare "*Tu-Io*", poiché, secondo il nostro autore, Pedro Laín non può che essere d'accordo che il principio relazionale della realtà umana è l'altra persona, ovvero, il *Tu* "*quale apriori della relazione*" personale.

Per comprendere meglio lo studio che questi fa sulla teoria di Pedro Laín e le figure di persona, prossimo e amico che rappresentano la realtà dell'altro nello spazio relazionale, è opportuno seguire l'ordine che l'autore ha dato all'opera.

Nel primo capitolo Bombaci realizza un veloce, ma essenziale, percorso biografico sulla figura lainiana, non soffermandosi più del dovuto sulle vicende accadute nel periodo falangista e franchista; inoltre questi presenta le teorie e gli autori che hanno avuto un forte impatto nella figura di Pedro Laín nello sviluppo della sua formazione intellettuale come medico e filosofico.

Dopo aver affrontato le vicende biografiche e alcuni argomenti propeudici alla narrazione, nel secondo capitolo l'autore ripercorre in maniera storiografica come la filosofia abbia scoperto la realtà dell'essere umano in quanto *altro* e soprattutto come la figura dell'*altro* sia stata studiata nella storia della filosofia spagnola da una prospettiva antropologica. In Spagna Pedro Laín, dichiara Bombaci, discute sulla figura personale dell'essere umano con i suoi principali interlocutori e maestri, Ortega y Gasset e Zubiri in un dibattito a distanza che riguarda varie discipline, "tra le quali l'antropologia filosofica assume la massima rilevanza, segnatamente nella declinazione propostane da Xavier Zubiri". Bombaci evidenzia che il dialogo fra Entralgo e Zubiri si incentra principalmente sulle opere antropologiche zubiriane; inoltre è assai interessante osservare in questo capitolo come il nostro autore ponga una forte attenzione alla discussione circa la realtà e la figura dell'altro, mostrando la concezione elaborata dai moderni e in particolare in particolare riguardo a Cartesio.

Tali discussioni continuano nel capitolo terzo, in cui sono posti in dialogo Pedro Laín e lo psicologo tedesco Theodor Lipps e con la sua teoria dell'empatia, intesa come una sinergia tra il soggetto esperiente e l'oggetto della esperienza. Lipps, secondo Bombaci, non arriva ad un riconoscimento dell'altro come *altro* in quanto *Tu* e questo porta l'autore spagnolo a rivolgere una critica radicale a tale teoria giacché Lipps stesso osserva l'altro come *alter ego* come realtà cosale, ovvero come realtà che è un corpo che appartiene ad un Io, e non come il "*Tu*" che porta all'*Io* all'orizzonte relazionale del "noi".

Fino adesso lo studioso di Pedro Laín ha svolto un ottimo e appropriato lavoro di inquadramento sia storiografico sia intellettuale riguar-

do alle vicende che ha avuto il filosofo aragonese per edificare la teoria dell'Altro.

Con questa premessa teorica si apre il capitolo quarto, Bombaci presenta l'*incontro personale* come quello che "mi riguarda (*me afecta*), mi tocca nell'interiorità, è realmente altro rispetto a me", poiché nella realtà dell'incontro l'*altro* è una persona diversa è un'altra realtà che apre alla realtà del noi. Molto interessante è l'affermazione di Bombaci riguardo all'incontro personale nella teoria lainiana, giacché essa è costituita dall'amore nella sua forma personale in quanto mi *tocca* nell'intimità. Qui il nostro autore attua una *fenomenologia dell'incontro* poiché in tale realtà "avverto dinanzi a me la presenza di un altro essere umano".

L'incontro per Pedro Laín, afferma Bombaci, non è soltanto un momento fisico anche se la corporeità "assume uno statuto privilegiato", giacché in essa si manifesta l'esistenza della persona come momento personale che "connota altresì ogni manifestazione empirica dell'uomo". Entralgo, utilizzando la terminologia zubiriana, afferma che l'incontro è apertura verso la realtà dell'altro e alla realtà delle cose perché l'essere dell'uomo, asserisce lo studioso lainiano, è *essere-con*, ecco perché "il presupposto primo dell'incontro è costituito dalla categoria ontologia della relazione". Continuando sulla scia di Zubiri il filosofo aragonese, in *Teoria y realidad del Otro*, offre quattro caratteri che sono fondamentali all'esistenza dell'essere umano, *genitivo*, *coesistenziale*, *dativo*, *comprensionale* ed *immaginario*. Questi caratteri sono attribuiti alla realtà della persona, ed è opportuno che essi siano considerati di grande importanza da parte del lettore, poiché in essi si trova la parte essenziale del pensiero relazionale lainiano. Il carattere genitivo è *coscienza di*, ovvero è anche *coesistenza*, intendendo co-esistere *con* l'altro; il carattere dativo è un *impulso ad essere*, vale a dire *tendere-a* un fine, in cui si manifesta come la vita sia *vivere per*, la quale porta a *co-esistere* con l'altro.

Il quinto capitolo è dedicato alla figura del maestro spagnolo Xavier Zubiri e all'influenza che il suo pensiero antropologico ha sulla teoria lainiana a riguardo dell'Altro. Bombaci presenta Zubiri come un filosofo esemplare che ha in sé una maniera profonda ed esigente nel pensare il problema dell'uomo; Zubiri infatti osserva come l'essenza dell'uomo non si risolve del tutto nella esistenza e per tanto la ricerca di tale fondamento deve essere rinvenuta altrove, ovvero, in ciò in cui si fonda l'esistenza, ossia la nuda realtà, poiché come afferma il filosofo basco la realtà non è "un attributo dell'Essere", bensì essa lo trascende e lo risolve in *realitas in essendo*.

La persona è una delle figure lainiane che Bombaci approfondisce nella sua opera; questi osserva che nella formazione del concetto e della realtà di persona, Pedro Laín, per spiegare la persona stessa, fa suo il concetto zubiriano di "*essere de suyo*", ossia che la persona sia quella figura che

possiede se stessa in quanto realtà, e per tale motivo si presenta all'altro nella forma dell'incontro. Continuando sulla scia zubiriana di persona, il nostro autore osserva che Laín nella costruzione della figura di persona utilizza anche il concetto delle *note* o di proprietà come quello che caratterizzano la persona in quanto tale. Queste proprietà non esauriscono la natura e l'essenza di ciò che la persona è, poiché essa è "non indifferente, non quantificabile, non esteriore".

Tale prospettiva apre il sesto e ultimo capitolo dedicato alle altre due figure rappresentate nella presenza e la realtà *inoggettivabile* della persona stessa, la figura del prossimo e dell'amico. È assai interessante come Bombaci per spiegare la figura del prossimo nella teoria lainiana, prenda come esempio la pericope evangelica del buon samaritano interpretata nei termini di incontro esemplare, giacché il samaritano non offre soltanto qualcosa che ha in più, bensì dona sé stesso e per tanto la sua intimità. In tale racconto Bombaci ravvisa l'espressione più elevata dell'amore che è *costante* e che diventa amore di *prossimità*, poiché per Pedro Laín, percepire l'altro in quanto persona "costituisce il prerequisito della prossimità". Notevole è l'indagine linguistica che Bombaci intraprende sulla parola utilizzata dal filosofo aragonese "*prójimo*", affermando che il prossimo è colui che sta accanto all'altro, che crede nella sua realtà personale e si dona all'altro manifestandosi in quanto altro che ama e che *con-sta* con l'altro.

Il nostro autore indaga l'ulteriore figura dell'amico, che per Laín è quella persona che si dona all'altro in vista del bene non soltanto suo, ma anche dell'altro e degli altri esseri umani. Per completare il ritratto di tale figura, Bombaci mette in evidenza, di nuovo, l'influenza zubiriana nella filosofia lainiana, osservando che l'amico è *sostantività*, come colui che si va costituendo e appropriando di sé sul piano convivenziale.

Un punto che appare evidente nell'opera e che non si deve lasciar inosservato, è la forte presenza della tradizione filosofica spagnola – ancora da indagare pienamente – con figure come Zambrano, Ortega, Zubiri, Rof Carballo; presenza che mostra in questa maniera l'importanza che questa stessa tradizione ha ancora oggi per il pensiero continentale.

Alla fine il lettore potrà osservare in questo studio come la figura dell'altro e dell'alterità sia sempre qualcuno che si presenta come persona, prossimo, amico manifestandosi nella forma personale dell'incontro. Tra questi concetti non esiste una gerarchia ma una necessità di relazione, poiché se l'amico è di fronte all'"*Io*" significa che è stato riconosciuto primariamente come persona e come qualcuno che è prossimo; ecco che si attua in questo studio una "fenomenologia dell'incontro" che inizia proprio dalla realtà relazionale del *Tu* e che porta alla realtà convivenziale del noi.

Roberto Di Ceglie

Francesco Paolo Adorno, *The Transhumanist Movement*, Palgrave Macmillan, Cham 2021, pp. 233

Questo volume di Francesco Paolo Adorno intende offrire un'introduzione al transumanesimo e al tempo stesso un'analisi critica di alcuni suoi aspetti di fondo. La partizione interna si adegua felicemente a questo fine. Dopo l'introduzione, seguono cinque capitoli che consentono al lettore di addentrarsi con sempre maggiore profondità e spirito critico nelle problematiche connesse al transumanesimo. Il primo capitolo si dedica a presentare i termini delle posizioni transumaniste, a cominciare dalla parola "*transhumanism*" e fino alla costellazione di movimenti e organizzazioni che della cultura transumanista si sono fatti promotori. Il secondo capitolo si interroga sulle ascendenze storiche del transumanesimo, illustrando quale radicamento se ne dia nella modernità umanistica e illuministica. Il terzo capitolo entra – per dir così – nel merito dei contenuti delle posizioni transumaniste, ed è infatti intitolato *Theory and Practice of Transhumanism*. Col quarto capitolo (*Problems of Transhumanism*) Adorno passa più decisamente a sottoporre quei contenuti a una rigorosa analisi critica; critica che tuttavia non manca mai nel corso dell'intero volume. Infine, nel quinto capitolo (*Transhumanism and Biopolitics*), l'autore offre una riflessione sul ruolo che le prospettive transumaniste svolgono nel contesto della vita politica della moderna società occidentale.

Come già anticipato, nel primo capitolo Adorno presenta una mappatura del transumanesimo, soffermandosi sul concetto e i suoi significati, sui protagonisti (inclusi i più o meno lontani precursori), sulle organizzazioni che ne promuovono le idee, sulle dichiarazioni che di quelle idee costituiscono un veicolo autorevole. Il secondo capitolo sembra costituire una sorta di riflessione intorno al concetto stesso di *trans-umanesimo*, giacché vi si indaga la connessione con l'umanesimo rinascimentale e con le radici di quest'ultimo, ovvero il valore classico della *humanitas*. In che senso il transumanesimo porta *oltre* l'umanesimo? Va inteso come sua continuazione e intensificazione oppure come suo tradimento e superamento nel *post-umanesimo*? Adorno sottolinea la continuità, giacché i transumanisti apprezzano e propagano la ricerca di libertà, democrazia, tolleranza e razionalità proprie dell'umanesimo. Sottolinea però anche la discontinu-

ità, che sembra caratterizzare più che altro i mezzi utilizzati, giacché all'educazione e alla riflessione dell'umanesimo il transumanesimo sostituisce le biotecnologie. Certo, anche intorno all'uso delle biotecnologie ci si può chiedere se il loro futuro risulti in continuità rispetto al passato oppure no. Si tratta cioè solo di incrementare le facoltà già presenti o di crearne di nuove? Adorno mostra che anche su questo la riflessione rimane aperta, come dimostrato dal fatto che a riguardo gli stessi transumanisti offrono convinzioni diversificate (cfr. p. 58). Fra tali convinzioni spicca quella di considerare "naturale" ogni proposta, per quanto innovativa essa sia. In tal senso, finanche i più futuristici ritrovati della tecnologia offrirebbero semplicemente il modo per meglio sviluppare quanto è richiesto da sempre, sin dal primo apparire dell'umano sulla terra (cfr. pp. 42-43). Ovviamente, in base a una tesi come questa, ogni critica del transumanesimo finirebbe per rivolgersi anche ai valori dell'umanesimo rinascimentale e moderno (cfr. p. 64). Ma a nostro parere potrebbe anche conseguirne una tesi diversa, ossia che il rifiuto dell'umanesimo si accompagni all'esaltazione del transumanesimo, nonostante quest'ultimo presenti come "naturali" le proprie proposte. Per spiegarci meglio, partiamo col considerare che Adorno si riferisce opportunamente all'insorgenza, prima ancora e autonomamente rispetto al transumanesimo, dell'anti-umanesimo di pensatori come Foucault e Althusser, di correnti di pensiero come il femminismo e i *race studies*. È un anti-umanesimo che mette sotto accusa l'idea dell'uomo come centro dell'universo, un'idea che implicherebbe un primato sugli altri animali e risulterebbe quindi portatrice di discriminazioni. Da ciò al post-umanesimo il passo è breve. Le nuove tecnologie, specialmente l'ibridazione con animali e ritrovati tecnologici, metterebbero fine all'idea di umano sostenuta dall'umanesimo, un'idea di umano che in realtà non avrebbe sufficiente fondamento. Ne segue allora la posizione che prima menzionavamo: possono convivere la tesi che il transumanesimo risulti in linea con esigenze presenti dagli albori della storia e quella per cui l'umanesimo, non rispondendo in realtà a tali esigenze, debba essere superato. Si rifiuta quindi l'umanesimo e si afferma il transumanesimo pur sostenendo la "naturalità" delle innovazioni di quest'ultimo.

Il capitolo terzo entra nel vivo delle problematiche implicate dal transumanesimo, e le affronta incentrando l'attenzione perlopiù sul concetto di *enhancement*. Adorno ne mette in luce i caratteri di fondo, gli interrogativi che si possono avanzare a suo riguardo e quindi i rischi che da esso possono seguire. Una volta definito (come è già stato fatto da altri, dicendo che esso è "ogni cambiamento biologico o psicologico che incrementi le possibilità di condurre una vita buona in un determinato insieme di circostanze", p. 74, trad. nostra, qui e d'ora innanzi), si potrebbe ritenere che si tratti di un concetto per nulla peculiare del transumanesimo, essendo di per sé parte dell'esperienza umana nel suo complesso (in tal

modo si propenderebbe per la continuità del transumanesimo rispetto alle modalità con le quali l'umanità ha tradizionalmente cercato di superare i propri limiti). Adorno mette però anche in evidenza che autori come Bostrom e Savulescu evitano definizioni dell'*enhancement*, preferendo procedere caso per caso (cfr. p. 69), segno che non risulta affatto semplice, perlomeno in determinate circostanze, capire se un certo intervento sulla natura umana sia oppure no *enhancement*. Potrebbe infatti consistere nel solo portare a compimento quello che già si è, ma anche nel *superare* ciò che si è. Si tratterebbe quindi di semplice *fulfilment* o di *hubris*? (cfr. pp. 72-73). Evidentemente, una risposta è possibile solo sulla base di una previa e valida definizione dell'umano. E fornirla non è affatto semplice, data la natura "politeistica" delle nostre società (cfr. p. 73).

Negli ultimi due capitoli, dedicati uno ai *problemi* del transumanesimo e l'altro al ruolo della biopolitica, al lettore sembrano offrirsi due piste di ricerca, che emergono proprio da quanto si è finora rilevato. Da un lato il tema amplissimo del rapporto fra terapia ed *enhancement*, dall'altro le problematiche che emergono quando si consideri la caratura sociopolitica della tecnologia contemporanea.

Circa il rapporto fra terapia ed *enhancement*, è evidente che esso riproponga ancora una volta la domanda sulla continuità o discontinuità del transumanesimo con quanto si è sempre fatto per superare limiti e difficoltà. Se il transumanista convinto della continuità ritiene di voler solo incrementare il proprio dell'umano e la sua dignità, i teorici della discontinuità ritengono invece che tale dignità venga messa in discussione da chi tende a superarne la natura di fondo. A detta di questi ultimi, insomma, l'*enhancement* costituirebbe un "perfezionamento non-terapeutico". Ma questo comporta la difficoltà di definire prima ancora che cosa sia "terapia" e ancora più fondamentalmente che cosa sia "malattia". In questo senso, Adorno cita Eric Juengst (cfr. p. 131), ad avviso del quale la definizione di *enhancement* in relazione alla terapia diviene un compito arduo giacché sempre relativo alle previe definizioni appena menzionate. Arduo appare poi anche il riferimento al concetto di dignità umana, che sarebbe vago (cfr. p. 123) e quindi bisognoso di precisazioni che, ancora una volta, la società politeistica prima evocata non può fornire senza che ciò, anziché mettere fine ai dibattiti, ne apra di ulteriori. In ogni caso, in questa complessità di dibattiti sembra esservi notevole spazio per un'assimilazione di terapia ed *enhancement*. Un'assimilazione sostenibile col ricorso alla fondamentale figura del *cyborg*, il cui ibridismo non risulta affatto sconvolgente se si considera che a consentirci di fare quello che intendiamo fare possono essere i neurotrasmettitori allo stesso modo degli occhiali. Si può quindi arrivare a dire che di *cyborg* si possono dare diverse specie (cfr. p. 164) e che la vita nel suo complesso risulta medicalizzata, giacché, per rimanere all'esempio appena riportato, si potrà

intervenire sulle funzioni cerebrali e allo stesso modo su quelle visive col fine di potenziarle entrambe, al di là di valutazioni meramente quantitative finalizzate a stabilire se e quando un certo intervento sia oppure no meramente terapeutico.

La stessa complessa articolazione di argomenti emersa finora emerge anche a proposito della tecnologia intesa come un insieme di strumenti *neutrali*, che toccherebbe a noi stabilire in che modo e con quali fini utilizzare. In realtà, come Adorno rileva, si tratta di strumenti tutt'altro che neutrali. E questo risulta in linea col fatto che di tali strumenti sono *altri* ad avere il controllo. Ne emerge quindi una contraddizione, per la quale la tecnologia transumanista vorrebbe metterci in grado di aumentare il controllo su noi stessi, ma allo stesso tempo ciò è realizzabile solo attraverso strumenti il controllo che sono *di altri* (cfr. p. 170). Secondo una sorta di eterogenesi dei fini, il transumanismo liberante finisce per condurre a un vero e proprio antiliberalismo (cfr. p. 178). La dimensione politica del transumanismo emerge quindi con prepotenza. Adorno mette più volte in guardia contro il riduzionismo materialistico, che porta a concentrare l'attenzione sulla sola dimensione biologica del corpo umano, dimenticando che esso è quello che è solo nella costante interazione con l'ambiente che lo circonda.

Facendo interagire la riflessione di pensatori come Arendt e Foucault, Adorno nota che, prima ancora della stagione transumanista, il concetto di lavoro si è posto al centro della riflessione politica riconducendo quest'ultima al primato dell'interesse economico. Il lavoro riguarda la produzione di beni come pure la trasformazione della natura e il ciclo della vita. Ponendo il lavoro al centro dell'attività politica, l'essere umano diviene necessitato, come ogni animale, dalla riproduzione e dal perfezionamento della vita (cfr. p. 183). Se a questo si aggiunge la medicalizzazione pervasiva sopramenzionata e la tesi di Foucault secondo cui si è andata affermando nel corso della modernità una vera e propria "polizia" sociale, sorta di tecnica di governo finalizzata a definire tutto quanto riguarda l'uomo e la sua felicità, ne segue che la politica non può più pensarsi se non come "biopolitica", laddove il corpo umano è soggetto alla politica e ne diviene campo di applicazione (cfr. p. 180). Opportunamente, Adorno ritiene che "il transumanismo e la polizia del Settecento siano più prossime l'uno all'altra di quanto si possa pensare in quanto non solo hanno fini simili – operare per il benessere degli individui – ma neanche i loro mezzi sono sostanzialmente diversi e sono essenzialmente, sebbene non esclusivamente, fondati sulla medicina" (p. 190).

In conclusione, Adorno offre una lettura del fenomeno transumanista tale da mostrarcene le numerose sfaccettature, i rischi e le contraddizioni. Se nel complesso il libro può apparire eccessivamente critico nei confronti degli aspetti che del transumanismo sono stati considerati, que-

sto va solo addebitato all'operazione meramente filosofica che l'autore propone. Essa si sostanzia di argomenti critici finalizzati al superamento di pregiudizi irragionevoli, siano essi a favore o a sfavore del transumanesimo. Certo, l'autore è consapevole che vadano operate delle scelte, e di queste si assume dall'inizio la responsabilità. Sin dall'introduzione, infatti, nota che nel volume intende affrontare solo alcune tematiche di fondo del transumanesimo, giacché per "movimento transumanista" va intesa una congerie troppo variegata di tendenze, gruppi, movimenti e organizzazioni (cfr. p. 1), con convergenze tra i vari gruppi che risultano spesso superficiali (cfr. p. 13). Una scelta della quale però l'autore non dà conto, e che invece avremmo apprezzato avesse perlomeno menzionato, è quella di non aver preso in considerazione il rapporto fra transumanesimo e trascendenza religiosa, che pure risulta notevolmente indagato nel contesto accademico di lingua inglese. Volumi collettanei quali *Transhumanism and Transcendence*, curato da R. Cole-Turner (Georgetown University Press, Washington DC 2011), *Religion and the Implications of Radical Life Extension*, curato da D.F. Maher e C. Mercer (Palgrave Macmillan, New York 2009), e ancora *Religion and Human Enhancement*, curato da C. Mercer e T. Trothen (Palgrave Macmillan, Basingstoke, Hampshire 2017) sono solo alcuni dei prodotti di una notevole discussione che impegna da tempo numerosi pensatori. Per non citare l'afflato religioso dei tanti che, in veste di precursori come pure di attori e di studiosi del transumanesimo, porta a ritenere non rifiutabile a priori l'espressione di Patrick Hopkins, ossia che il transumanesimo, sebbene non sia una religione, è tuttavia "religioso" (*Transcending the Animal: How Transhumanism and Religion Are and Are Not Alike*, in "Journal of Evolution and Technology", n. 14 2005, p. 22). È vero che transumanisti di spicco come Max More marcano il carattere secolare del transumanesimo affermando come una mera filosofia di vita. Ma proprio la varietà di posizioni interne al transumanesimo e la corrispondente consapevolezza da parte di Adorno di aver dovuto operare delle scelte avrebbe richiesto a nostro parere una previa spiegazione delle ragioni per non considerare un'area di interesse come quella in questione nell'ambito di un'introduzione al transumanesimo. Tutto questo, però, toglie solo nella misura in cui si consideri quanto Adorno *non* ha fatto. Per quello che però ha fatto, e che costituisce il libro in questione, è bene ribadire, semmai ve ne fosse ancora bisogno, che la numerosità degli argomenti e dei dibattiti considerati, e la loro valutazione non pregiudiziale, fanno di questo volume un'ottima introduzione al pensiero e alle pratiche del transumanesimo.

Michele Farisco

Ortensio Zecchino, *Perché non possiamo non dirci “cristiani”*.
Lecture e dispute sul celebre saggio di Benedetto Croce,
Rubbettino Editore, Saveria Mannelli 2024, pp. 254

Tra i giganti della cultura italiana, Benedetto Croce occupa indubbiamente una posizione di assoluto rilievo. A prescindere dalla sintonia o divergenza rispetto al suo pensiero, chiunque sia interessato a ricostruire le principali linee di sviluppo della cultura italiana ed europea degli ultimi due secoli non può prescindere da un confronto con l'architettura teorica elaborata da colui che è unanimemente considerato tra i maggiori protagonisti di entrambe. La poliedricità della riflessione e dell'attività di Croce è tale da non potersi limitare all'ambito filosofico, in particolare agli apporti da lui forniti al liberalismo e al neoidealismo, che pure contribuì a fondare e sviluppare quale principale teorico di entrambi. I campi nei quali l'impatto di Croce è stato, e ancora è significativo si estendono alla politica, sia come esercizio di responsabilità di governo sia come elaborazione di una dottrina, alla storia, che per Croce è l'ambito privilegiato della ragione, alla religione, pensata in una irrisolta dialettica tra razionalità dello spirito e religiosità tradizionale e confessionale. Il famoso scritto *Perché non possiamo non dirci “cristiani”*, pubblicato su *La Critica* il 20 novembre del 1942, riflette pienamente la complessità dell'opera crociana.

Evidenziare una tale poliedricità di interessi e ambiti d'azione del pensatore nato in Abruzzo ma vissuto prevalentemente a Napoli è uno dei meriti principali del presente saggio di Ortensio Zecchino, che in verità ne presenta anche molti altri, non ultimo una ricostruzione analitica molto dettagliata della genesi e delle risonanze, entrambe alquanto travagliate, del saggio di Croce. Tale ricostruzione è fondata sulla consultazione diretta di fonti tutte rilevanti e talvolta inedite (si pensi alla corrispondenza con Maria Curtopassi, con Giuseppe De Luca, con Guido Gonella, con Monsignor Olgiati, e soprattutto con Luigi Sturzo, del quale si riporta un poco conosciuto commento al saggio crociano uscito in lingua inglese il 15 luglio 1943), tra cui illuminanti passi dal *Diario* personale di Croce. L'acutezza ermeneutica di Zecchino è davvero ammirevole, nutrita da numerosissimi riferimenti agli scritti crociani precedenti e successivi il *Perché non possiamo non dirci “cristiani”*, nonché dalla lettura della ric-

chissima critica filosofica del pensiero di Croce. Ne viene fuori un saggio esso stesso poliedrico e complesso come il suo oggetto di studio, pur prevalendo una chiave ermeneutica di tipo storico, un approccio quanto mai pertinente, che rispecchia un tratto saliente della filosofia, pur idealistica, di Croce, ossia il suo legame con l'attualità politica del tempo, che ha voluto elevare alla forma del pensiero senza con questo risolversi in astrattismo speculativo.

In questo senso, nel libro si evidenzia giustamente la sofferenza "spirituale" degli anni della Guerra quale genesi primaria dello scritto di Croce: la concezione cristiana della vita è presentata dal filosofo quale argine e rimedio alla concezione pre-cristiana, pre-ellenica e pre-orientale, ossia alla "barbarica violenza dell'orda", che stava prevalendo agli inizi degli anni '40. Croce è ben consapevole della portata della crisi del suo periodo storico, ossia il rischio della fine della civiltà (europea). La storia lo conduce a sentimenti di desolazione e finanche pessimismo per un crescente "scivolamento nella barbarie", in apparente contrasto rispetto alla prospettiva storicistica e alla concezione dialettica crociana (emblematiche in tal senso *La fine della civiltà* e *L'Anticristo che è in noi*, analizzati nel capitolo III), infine apparentemente recuperate nella riaffermata "coincidenza dello spettacolo della storia con la verità dell'etica". Tuttavia, come evidenziato da Zecchino, gli eventi storici coevi al Croce testimoniano un "generale abbassamento del sentire morale, individuale e collettivo". Filosoficamente, Croce individua la genesi del male in quella che egli definisce la quarta forma dello spirito accanto al Vero, al Buono e al Bello: la Vitalità. Questa è a un tempo condizione necessaria per l'effettiva realizzazione storica delle altre tre forme, e costante minaccia a tale realizzazione. Le vicende belliche della prima metà del XX secolo dimostrano, infatti, che il male radicato sulla Vitalità può rompere la legge del progresso storico verso la libertà, ossia incrinare le certezze filosofiche storicistiche e dialettiche. Certezze infine riaffermate da Croce, in un non casuale stretto dialogo con Hegel, rispetto al cui sistema non rinuncia ad affermare significative riserve, puntualmente evidenziate da Zecchino.

Come noto, nel saggio *Perché non possiamo non dirci "cristiani"* Croce indica nel cristianesimo un solido appiglio per arginare la deriva di cui sopra. Giustamente Zecchino sottolinea che tale richiamo al cristianesimo ha una valenza anzitutto politica piuttosto che confessionale: la rivoluzione del cristianesimo è stata una rivoluzione di civiltà, o per meglio dire morale, che ha posto le basi per lo stesso stato liberale moderno. La peculiarità della rivoluzione cristiana consiste nel suo rivolgersi all'interiorità della coscienza morale. Che tale richiamo al cristianesimo da parte di Croce non si risolva in adesione confessionale ha un solido fondamento teoretico nella sua concezione storicistica, secondo cui la storia è allo stesso tempo sviluppo e pienezza dello spirito, e la forma di conoscen-

za più elevata è hegelianamente la filosofia quale logica della Ragione (*Vernunft*) contrapposta all’Intelletto (*Verstand*): solo la prima può comprendere Dio come Spirito, una comprensione che in senso stretto non è possibile neppure alla religione rivelata (che infatti, come ampiamente evidenziato da Zecchino, viene vista da Croce perlopiù come *preparazione* del sapere filosofico). Da ciò il contrasto culturale con la teologia cattolica coeva e anche successiva, come dimostrato dall’editoriale de *La Civiltà Cattolica* richiamato nel libro (che pure ha motivazioni di tipo politico, oltretutto filosofico).

Lo stesso richiamo del Croce al cristianesimo quale necessario riferimento per la salvaguardia della civiltà pone un primo interrogativo, che rimane sullo sfondo del libro di Zecchino, che pure contiene rilevanti spunti per una più elaborata risposta: cosa si intende precisamente per civiltà? Croce scrive da occidentale e teorico del liberalismo, per cui è lecito concludere che questa sia la civiltà che egli vuole salvaguardare. Traslando la sua analisi nel contesto contemporaneo, tentandone una non esente da rischi attualizzazione, emergono due questioni correlate: (1) nel mondo globalizzato di oggi, identificare la civiltà con i valori occidentali e liberali non è scontato, e deve giocoforza confrontarsi con la dialettica di valori altri (ossia con altre civiltà) che entrano prepotentemente nelle dinamiche culturali e civili dell’occidente stesso (come evidenziato nella seconda parte del libro, in particolare nel capitolo VI, *Il saggio nel tempo nuovo della cattolicità*). Partire dal presupposto che la civiltà si identifichi con i valori dell’occidente rischia, pertanto, di essere inefficace nella pur necessaria mediazione con civiltà e valori altri. (2) L’identificazione del cristianesimo con la morale cristiana, che surrettiziamente sembra sorreggere il discorso crociano, vale fino a un certo punto. Sul piano filosofico, la riduzione del cristianesimo a legge morale viene rimproverata da Nietzsche a san Paolo, mentre il nocciolo della predicazione di Gesù è il superamento della legge stessa. Tuttavia è vero che il cristianesimo, soprattutto quello confessionale, si è di sovente ridotto a precettistica e quindi a regolamentazione dell’agire (a “legalismo”, per usare un termine crociano), e che tale visione ha condizionato e ancora condiziona la vita, individuale e sociale, di milioni di persone. Ma, almeno sul piano teorico e dogmatico, l’eccedenza del cristianesimo rispetto alla morale mantiene la sua importanza.

Un altro punto teoreticamente importante, nonché collegato al precedente, è il presupposto storicistico dell’opera di Croce, che pure segna una distanza significativa con la teologia cattolica, la quale, non a caso, lo critica aspramente su questo aspetto. Quando Croce parla della “coincidenza dello spettacolo della storia con la verità dell’etica” non può non entrare in cortocircuito con la prospettiva di una religione rivelata, per cui la storia è sospesa tra “il già e il non ancora” dell’*eschaton*, che ha

avuto un preciso *incipit* storico (l'incarnazione) ma attende la definitiva realizzazione (la seconda, definitiva venuta del Verbo incarnato). La mia impressione (ma trattasi soltanto di un'intuizione che andrebbe suffragata da ulteriori, più puntuali verifiche) è che Croce, radicato in un irrinunciabile immanentismo, riduca il cristianesimo a morale per neutralizzarne la "sporgenza" metastorica e trascendente (mi pare che la lettura di Del Noce puntualmente riportata nel testo di Zecchino vada in questa stessa direzione). Su questo punto, in realtà, lo stesso Croce presenta un atteggiamento potenzialmente contraddittorio: se da un lato lo storicismo assoluto nega la trascendenza, dall'altro emerge, soprattutto dalle pagine del *Diario* riportate da Zecchino, uno stato d'animo di Croce come "ostinatamente fermo nel suo immanentismo e pure aperto alla trascendenza". Una delle tante apparentemente irrisolte contraddizioni del pensiero crociano evidenziate nel libro, che ne fanno un pensiero vivente.

Un altro interrogativo di tipo ermeneutico sorge in riferimento alla categoria della Vitalità, che Croce afferma doversi domare da parte della "sempre risorgente libertà". Mi chiedo se ci sia un riferimento implicito alla *Lebensphilosophie* del XIX secolo, con cui Croce condivideva la vocazione antipositivistica e la visione dell'esistenza come tensione del finito verso l'infinito, ma non l'esito irrazionalistico ed esistenzialistico *ante litteram*. In ogni caso, mi sembra che la vitalità sia un elemento necessario nella concezione dialettica di Croce, per cui non può darsi libertà senza il suo opposto: questo punto richiama alla mente la concezione scheleriana del *Geist*, che è di per sé impotente e capace di agire ed esprimersi storicamente solo se sollecitato e "provocato" dall'esterno.

Un altro punto che il libro di Zecchino meritoriamente introduce e che richiederebbe ulteriori approfondimenti è la concezione di libertà che sorregge il testo di Croce, che segna una distanza critica importante con la teologia cattolica, in rapporto alle declinazioni della libertà che soggiacciono all'interazione tra sapere scientifico e tecnologico da un lato e società contemporanea dall'altro. In particolare, richiede ulteriori approfondimenti la visione della libertà come assoluto che è alla base del liberalismo di Croce, in relazione alla proliferazione dei (presunti) nuovi diritti cui Zecchino accenna nelle pagine finali del libro a partire dal dialogo tra Habermas e Ratzinger. La mia impressione è che tale tendenza contemporanea sorga da una concezione equivoca della libertà come assoluto.

Tale equivoco si collega, tra l'altro, agli scenari contemporanei, cui Zecchino fa cenno nel già citato capitolo VI, segnati dal contrasto tra secolarizzazione e nuovi fondamentalismi religiosi. Uno dei rischi connessi a tale contrasto è un nuovo materialismo centrato su un edonismo individualistico elevato a diritto, risultato dell'assolutizzazione della libertà di cui sopra. In aggiunta a questo rischio ne emergono altri due,

uno più subdolo e uno più prettamente politico. Il primo è la minaccia dell'anonimia, ossia della perdita di identità individuale nella massificazione delle opinioni e dei valori, ossia nel mondo della "post-verità" dei social-media. Il secondo, connesso al primo, è la minaccia dei nuovi poteri finanziari e politici, che si auto-alimentano in un circuito auto-catalitico per cui più si utilizzano i loro prodotti (per esempio, i social media) più ne aumenta il bisogno (si pensi alle nuove patologie di dipendenza da social media). La combinazione di questi due pericoli genera nuove forme di disegualianza sociale e marginalità politica (per esempio nella forma dell'analfabetismo digitale, che di fatto preclude l'accesso a nuove forme di attività politiche), che credo siano in gran parte inedite nella storia. Il punto è, allora, se e come il messaggio di Croce nel *Perché non possiamo non dirci "cristiani"* possa offrire spunti rilevanti per fronteggiare queste nuove minacce. Basta richiamarsi alla rivoluzione morale del cristianesimo? O è oggi necessario integrarla con altri riferimenti culturali e civili necessari per accomodare nuove e inedite sfide politiche (*in primis* l'integrazione di prospettive religiose tendenzialmente anti-liberali)? Il saggio di Croce in oggetto offre spunti di riflessione rilevanti per rispondere a tali domane, rivelandosi particolarmente attuale per comprendere le dinamiche storico-culturali del tempo presente.

L'attualità del pensiero di Croce è confermata, tra l'altro, dal grande interesse espresso qualche anno fa dal mondo anglosassone, che ha incluso Croce tra i più grandi pensatori liberali del 900. Il saggio di Zecchino si può indubbiamente annoverare tra le letture più profonde ed equilibrate dello scritto crociano, capace come pochi altri di penetrare la profondità delle sue numerose diramazioni, autobiografiche, filosofiche, politiche, religiose, storiche, nonché la trama di rapporti con gli eventi e le relazioni che lo hanno preceduto e che lo hanno seguito. Uno strumento prezioso per chi voglia affacciarsi sulla complessità di un pensatore unico che ha attraversato da protagonista la storia del nostro Paese. Dal libro di Zecchino emerge, infine, un'intonazione non intellettualistica, il che evita il rischio di una lettura "distaccata" del pensiero crociano: di questo si evidenziano puntualmente i risvolti personali e, potremmo dire, esistenziali, *in primis* le molte contraddizioni irrisolte e la vocazione a un "credo dubitativo". Tra le righe si percepisce una sorta di sintonia o simpatia tra Zecchino e Croce, due spiriti estremamente sensibili ai drammi della storia. Tale rapporto di affinità travalica l'ambito intellettuale per toccare le corde più profonde dell'esistenza personale. Al di là dei meriti culturali di cui sopra, questo è forse l'aspetto più toccante e coinvolgente del presente saggio.

Adriano Lotito

Roberto Mordacci, *Critica e Utopia. Da Kant a Francoforte*, Castelveccchi, Roma 2023, pp. 185

Numerosi e partecipati sono i dibattiti che attraversano la teoria critica di ispirazione francofortese, a testimonianza di un durevole interesse nei confronti delle sfide che essa si è posta. Una prima sfida riguarda la fondazione normativa della critica, e in particolare il problema di come dedurre dei criteri che non siano astratti rispetto alle pratiche cui dovranno applicarsi, ma piuttosto immanenti alle stesse. La riflessione metateorica ha una decisiva ricaduta pratica: a seconda delle modalità con cui si attinge il metro del giudizio, si avranno strategie diverse per orientare l'azione contro l'ingiustizia. Qui emerge una seconda, incombente sfida, quella cioè di riannodare un legame tra teoria e prassi, facendosi carico di quell'interesse per l'emancipazione che è al cuore dell'impresa teorico-critica. Il libro di Roberto Mordacci affronta sapientemente queste sfide, fornendo indicazioni importanti su entrambi i fronti. La prima parte è dedicata a una classificazione originale delle forme della critica francofortese, mentre nella seconda parte si elabora un nuovo modello, la critica utopica, con l'intenzione di sopperire alla "limitata capacità propositiva" (p. 4) riscontrata negli approcci esistenti.

Mordacci individua quattro modelli che ricorrono nella tradizione francofortese, ai quali sono dedicati i primi quattro capitoli. Il primo è quello della critica trascendentale di matrice kantiana, di cui Mordacci tratteggia tre caratteristiche fondamentali: il rimando a un orizzonte universalistico formale, il raccordo con una filosofia teleologica non deterministica, e l'elaborazione di una prospettiva cosmopolitica (p. 28). La critica opera su due livelli: l'individuazione delle condizioni di possibilità interne a ciascuna pratica sociale, e il rimando a una struttura trascendentale della prassi, identificata con le condizioni del volere razionale, dunque libero. La giuntura è nel concetto di non-contraddittorietà del volere, filo conduttore dell'intero volume: non si può volere, a un tempo e sotto il medesimo rispetto, una cosa e il suo contrario. Al livello trascendentale, questo significa che il volere non può non volersi, cioè non volere la propria autonomia. Da cui un criterio capace di trascendere la contingenza storica. Tuttavia questo non fa della critica trascendentale una

critica esterna o trascendente, precisa opportunamente Mordacci contro un'interpretazione diffusa (per esempio R. Jaeggi, *Critica delle forme di vita*, Mimesis, Milano 2021). Il criterio normativo è infatti presupposto all'interno di ogni pratica. Inoltre, esso si traduce in una tendenza storica verso la libertà, incarnata nel telos cosmopolitico, senza che però vi sia alcuna direzione garantita. Mordacci mostra come il modello kantiano sia capace anche di superare il riduzionismo comunicativo della variante habermasiana. "L'aspetto normativo, proprio perché formale, riguarda tutte le forme dell'agire" (p. 40), compreso l'agire economico. Questa forma di critica mostra dunque di avere un grande potenziale, a patto che venga integrata con una indagine delle strutture di dominio sociale.

Il secondo modello è la critica dialettica, riconducibile a Hegel e Marx, e che Mordacci riconosce essere la principale forma di critica francofortese. L'elemento centrale è la contraddizione: si mette in luce il processo per cui i principi di un sistema sociale si rovesciano nel loro contrario in forza di se stessi, cioè a causa del loro fissarsi unilateralmente. Ciò determina la necessità di un superamento (*Aufhebung*) che riguarda ad un tempo il sistema sociale e le sue norme e che conduce a una reinterpretazione dei principi di partenza a un livello di maggiore universalità e concretezza. Il merito di questo modello è individuato da Mordacci nel nesso stabilito tra contraddizioni sociali e movimento trasformativo, racchiuso nella categoria di negazione determinata: il rovesciamento di una situazione contraddittoria può produrre i propri contenuti concreti solo lavorando con e dentro le contraddizioni di partenza e non sovrapponendovi dei valori sostanziali attinti altrove. Tuttavia l'assenza della dimensione trascendentale si traduce in un radicale storicismo, tale per cui si concepisce la critica in base a "criteri che, per definizione, non possono trascendere l'esistente" (p. 69). Per ovviare a questa chiusura, la critica dialettica ha introdotto una insostenibile teleologia necessitarista che fonda il trascendimento del dato storico su una logica oggettiva presupposta, sia essa il dispiegamento dello spirito o la dialettica dei modi di produzione. Mordacci rileva giustamente come la reinterpretazione della critica dialettica offerta da Honneth, soprattutto in testi più recenti come *L'idea di socialismo*, consenta di superare lo storicismo idealistico, in una direzione che riapre la critica a un orizzonte utopico, sebbene anche Honneth non disponga di un criterio ultimo, cioè trascendentale, della libertà.

Una terza forma di critica che Mordacci mette a tema è quella genealogica, ispirata a Nietzsche e Foucault e che, nella tradizione francofortese, è stata recepita in particolare in *Dialettica dell'illuminismo* (1947). Essa svolge un'analisi storico-interpretativa delle pratiche costitutive di una forma di vita, smascherandone il carattere originariamente repressivo, e fungendo così da premessa per un processo di liberazione. Anche nel-

la critica nietzschiana della morale emerge il tema della contraddizione pratica: la volontà della morale è “una volontà che mira ad affermarsi, negando se stessa, contraddicendo nell’azione il proprio stesso agire” (p. 76). Tuttavia l’esito, rileva Mordacci seguendo Habermas, sembra essere quello di un ritorno nostalgico alle origini. In Adorno e Horkheimer a essere decostruita è l’intera civiltà occidentale il cui atto originario è nell’auto-imposizione di una razionalità strumentale che stabilisce un rapporto di dominazione tra un soggetto calcolante e la propria natura interna ed esterna, in nome di una promessa di emancipazione che si capovolge nel suo opposto. Tuttavia non è chiaro, segnala Mordacci, se e come si possa uscire da questa contraddizione. Il rischio è nell’assunzione di una postura negativistica che paralizza la prassi. Una terza variante è offerta da Foucault. Si tratta di mostrare le pratiche da cui è sorto, senza alcuna linearità, un certo modello di soggetto, in contiguità con determinate idee e attraverso forme di disciplinamento dei corpi, così da smascherarne la contingenza e rendere accessibili opzioni alternative di soggettivazione. Mordacci rimarca l’importanza di una genealogia come “critica ironico-dissociativa-sacrificale” (p. 88) capace di sottrarre il senso storico agli usi della tradizione. Allo stesso tempo è rilevata l’incapacità foucaultiana di articolare un discorso etico come fondamento della diagnosi e di un progetto emancipativo.

L’ultima forma di critica viene definita “messianica”. Qui il punto di vista è quello di un’alterità trascendente in nome della quale si annuncia la possibilità di un rivolgimento rivoluzionario. L’emancipazione viene pensata come redenzione, in discontinuità con qualsiasi gradualismo storico. Mordacci attribuisce l’uso di questo metodo anzitutto a Walter Benjamin. Se l’invito a “spazzolare la storia contropelo” (W. Benjamin, *Sul concetto di storia*, Einaudi 2010, p. 31; Mordacci p. 95) è una doverosa cautela rispetto a ogni trionfalismo progressista, vi sono però due limiti segnalati da Mordacci. Il primo è l’impossibilità di ogni azione politica, conseguita al dichiarare imperscrutabile l’istanza emancipativa. Il secondo è la riduzione della storia al ritorno di un sempre-uguale privo di differenze di valore, come risultato dell’indeducibilità di un criterio immanente. Piuttosto che accantonare il modello messianico, la mossa acuta di Mordacci è di reinterpretarlo in chiave trascendentale, declinando il concetto benjaminiano di adesso (*Jetztzeit*) come trascendenza normativa immanente a ogni momento storico e dunque possibilità di giustizia sempre attivabile, in modi imprevedibili. Un’ulteriore variante della forma messianica è individuata da Mordacci nell’uso che Marcuse fa del concetto di utopia. In particolare nel *Saggio sulla liberazione* (1969), Marcuse interpreta l’utopia come possibilità reale, le cui condizioni di realizzazione risiedono anche nello sviluppo tecnologico, a patto che questo venga inquadrato in sistema sociale alternativo e si accompa-

gni alla negazione determinata dei falsi bisogni indotti dal capitalismo. Mordacci osserva che i veri bisogni a cui Marcuse allude non sono basati su una antropologia naturalistica, come spesso viene frainteso, ma sono il risultato di un rovesciamento delle contraddizioni date. Un'operazione che riveste un'importanza decisiva nel metodo utopico.

Il quinto capitolo è dedicato a un'utile ricognizione della teoria critica contemporanea, nella quale si rileva una maggiore propositività. Gli autori trattati sono Hartmut Rosa, Rahel Jaeggi, Rainer Forst e Amy Allen. Mordacci si sofferma con attenzione in particolare sulla riflessione di Jaeggi, mostrando molto appropriatamente come l'insistenza sull'elemento hegeliano coesista con un nucleo trascendentale non del tutto tematizzato ma senza il quale l'intero discorso di Jaeggi finirebbe per contraddirsi. La riformulazione del concetto di alienazione proposta da Jaeggi presuppone infatti un principio di "libertà positiva" (p. 115), definita come la possibilità da parte del soggetto di appropriarsi della relazione con sé, con il mondo e con gli altri. In un rapporto alienato il soggetto, non potendo esercitare una presa sul proprio sé e sulle interazioni sociali, finisce per volere ciò che non vuole e per non volere ciò che vuole, sperimentando una lacerazione della propria identità. Alla base dell'alienazione Mordacci scorge dunque nuovamente il problema della contraddizione pratica, non sviluppato da Jaeggi. La questione ritorna sul terreno della critica delle forme di vita. Di essa Mordacci valorizza in modo puntuale il carattere trasformativo, aspetto fondamentale del modello utopico. Tuttavia una critica immanente può tradursi in un processo di trasformazione radicale che sia normativamente giustificato, soltanto facendo appello a una struttura dell'agire che non sia del tutto riducibile alla forma di vita storicamente data, pur essendo costitutiva di ogni sua pratica.

L'esigenza di tenere assieme la dimensione storica e quella trascendentale, assieme all'urgenza di articolare una efficace *pars costruens*, motivano l'elaborazione di un ulteriore modello di critica. Nei capitoli sei e sette che costituiscono la seconda parte del volume, Mordacci sviluppa così le coordinate di una forma utopica della critica che debba affiancarsi e non sostituirsi ai modelli precedenti. Attingendo anche a lavori precedenti (si veda R. Mordacci, *Ritorno a Utopia*, Laterza, Roma-Bari 2020), si procede a una confutazione delle obiezioni rivolte al pensiero utopico, adoperando *Utopia* di Thomas More come punto di riferimento. Il racconto utopico non pretende di dedurre un modello ideale di società su basi metafisiche, bensì è una pratica immaginativa situata e realistica. Esso non astrae dall'analisi delle strutture sociali, bensì vi si fonda: il libro II di *Utopia*, rimarca Mordacci, è guidato dalla diagnosi delle contraddizioni della società inglese contemporanea a More, svolta nel libro I. Inoltre l'utopia parte dalle istanze della classe oppressa per immaginare la società

alternativa e, in More a differenza che in Bacone e Campanella, rifugge tentazioni autoritarie. Accanto all'immaginazione, Mordacci individua, come condizione del pensiero utopico, cinque "strutture fondamentali del senso comune" (p. 157): il senso di giustizia, la tensione verso l'eguaglianza, il desiderio di libertà, l'impulso alla solidarietà, l'aspirazione alla felicità. Queste capacità, frutto di una sensibilità morale condivisa, motivano l'immaginazione al cospetto di condizioni sociali percepite come intollerabili, sebbene esse non siano in grado di darci alcuna regola per valutare le prassi. Tale criterio, operante sia nel lavoro immaginativo sia nella sensibilità morale, viene identificato da Mordacci nel principio del rispetto, inteso come la regola della non-contraddizione del volere: "il potere della volontà che si riduce a mero mezzo contraddice l'originario potere affermativo del volere stesso" (p. 171). Il rispetto è dunque l'equilibrio delle libertà che rinunciano a dominarsi e decidono di accordarsi secondo ragioni. Si tratta di una regola immanente alla struttura della prassi in quanto tale, e che comanda il riconoscimento dell'umanità di ciascuno e dell'appartenenza a un comune regno di fini. Mordacci precisa il carattere formale di questo principio: esso non determina le immagini concrete dell'utopia, essendo queste elaborate da una immaginazione politica il cui esercizio è sempre situato e connesso con le contraddizioni pratiche che emergono in una determinata relazione tra poteri. La situazionalità della pratica immaginativa, specie in una società complessa, è ciò che spinge Mordacci a preferire un'utopia "frammentaria" piuttosto che "sistemica" (p. 174). Le contraddizioni globali potranno essere affrontate solo per coordinazione e connessione tra "realizzazioni utopiche locali". Sintetizzando questa forma originale di critica, Mordacci precisa quattro passaggi metodologici: l'analisi della realtà sociale, l'individuazione delle contraddizioni, l'immaginazione di assetti alternativi e la progettazione di pratiche trasformative. Un tale modello risulta compatibile con le forme di critica analizzate e al contempo ne costituisce un'integrazione necessaria, in virtù del ruolo centrale che in essa assume il lavoro immaginativo in opposizione alle contraddizioni rilevate.

L'organicità della proposta di Mordacci, il rigore argomentativo e l'ampio ventaglio dei riferimenti, nonché la chiarezza ed eleganza dell'esposizione, ne fanno un libro importante, che arricchisce notevolmente l'armamentario della teoria critica contemporanea, il cui pluralismo è giustamente considerato dall'autore una risorsa piuttosto che una debolezza. Volendo avanzare qualche osservazione, si potrebbero considerare due punti. Il primo riguarda il problema del punto di vista dei teorici e delle teoriche critiche. Qual è il posizionamento riflessivo della teoria critica, e dunque di coloro che la esercitano, nei processi sociali analizzati e giudicati, nel rapporto con gli attori che vi partecipano e nella produzione delle immagini utopiche? Da un lato, Mordacci assegna alla critica

“la funzione di rendere la contraddizione evidente e di segnalare la volontà coinvolte, in particolare a chi ne subisce gli effetti negativi, ma non solo”, per poi “immaginare scenari alternativi che rovesciano precisamente queste contraddizioni” (p. 168). L’aspetto diagnostico segue la direzione della critica dell’ideologia, con lo scopo di rischiarare i blocchi che occultano la contraddizione e inibiscono le volontà. Dall’altro lato, Mordacci prende le distanze dall’elitismo adorniano (p. 81) e rimarca il carattere democratico del pensiero utopistico, alieno sia a saperi esoterici sia al primato di competenze tecnocratiche (p. 147). La sfida allora è nell’articolazione di un dialogo tra un’analisi delle contraddizioni basata su competenze specialistiche, e una prospettiva centrata sulle prassi critiche, anche teoriche, agite dagli attori sociali in virtù delle proprie esperienze e specificità epistemiche (si veda in particolare R. Celikates, *Die Macht der Kritik. Epistemische Asymmetrien, Standpunkte und migrantische Praktiken*, in “WestEnd: Neue Zeitschrift für Sozialforschung”, 2, 2020, pp. 81-96). Questi punti di vista *subalterni*, lungi dal dover essere accettati acriticamente, potrebbero arricchire la diagnosi sociale, portando alla luce quei blocchi che distorcono lo sguardo degli stessi critici di professione, e allo stesso tempo, contribuire all’immaginazione e progettazione di pratiche alternative. Mappare le soggettività critiche e le loro iniziative di resistenza, ricostruirne la grammatica morale e le categorie epistemiche, valorizzarne l’immaginazione normativa, potrebbero essere direzioni stimolanti nel quadro di una critica utopica capace di interessare una relazione di mutuo (e polemico) scambio tra teorici specialisti e movimenti sociali.

Un secondo punto riguarda lo statuto dell’immaginazione. Mordacci individua in essa “la facoltà essenziale” (p. 148) del pensiero utopico, distinguendosi rispetto a trattazioni incentrate puramente sull’elemento normativo razionale. L’operazione di Mordacci consente, cioè, di evitare la trappola idealistica di un’autosufficienza del principio normativo, rispetto ad altri autori che hanno sovradimensionato la potenza della normatività, derubricando quelle dimensioni affettive e, appunto, immaginative, indispensabili alla traduzione dell’istanza critica in pratiche trasformative concrete. Il modello della critica utopica ha il potenziale di fare da raccordo tra le correnti più razionalistiche della teoria critica e quell’insieme di approcci letterario-dischiudenti che enfatizzano la dimensione estetica della normatività e le *narrazioni* articolate dagli attori sociali (per esempio M.P. Lara, *Moral Textures. Feminist Narratives in the Public Sphere*, University of California Press, Oakland 1998). Si potrebbe allora dare maggiore spazio alla funzione dell’immaginazione politica come condizione *poietica* della realizzazione della libertà nella storia e come dispositivo operativo della trascendenza nell’immanenza. Tale indirizzo potrebbe essere sviluppato anche in una direzione onto-

logico-sociale, dando risalto all'immaginazione come fattore costituente dell'azione sociale e un motore, non l'unico, del processo storico. Questo consentirebbe anche di riavviare un dialogo con l'opera di Cornelius Castoriadis, dopo le interlocuzioni che con essa hanno stabilito Habermas e Honneth. Un approfondimento di questi aspetti sarebbe interessante e potrebbe forse rafforzare ancora di più la peculiare fisionomia della proposta di Mordacci nel quadro della teoria critica contemporanea.

Angelo Tumminelli

Simona Tiribelli, *Identità personale e algoritmi*.

Una questione di filosofia morale,

Carocci, Roma 2023, pp. 131

Di fronte allo scenario sempre più complesso delle società tecnologiche abitate dai sistemi algoritmici, Simona Tiribelli, docente di Etica dell'intelligenza artificiale e Global Justice and Technology presso l'Università di Macerata, propone una riflessione filosofica sul rapporto tra identità personale e algoritmi a partire da una prospettiva etica. Nel suo recente volume, intitolato *Identità personale e algoritmi. Una questione di filosofia morale*, la studiosa si chiede cosa significa essere "persone" nelle società algoritmiche e in che misura tali sistemi influiscono sui processi di formazione delle identità personali e sull'esercizio della libertà di scelta. Analizzando in modo rigoroso e supportato dalle più recenti ricerche teoriche il contesto dell'ambiente *phygital*, quello caratterizzato da una imponente ibridazione tra spazio fisico e spazio virtuale, Tiribelli da un lato riconosce le enormi potenzialità connesse all'infosfera (cfr. Floridi 2019), la possibilità di stabilire connessioni prima inedite e di generare nuove modalità relazionali, e tuttavia non nasconde la preoccupazione propriamente etica legata allo sviluppo delle identità personali nel contesto di una sempre maggiore datificazione della realtà. Se, infatti, per un verso i sistemi algoritmici, quelli di tipo deterministico ma ancor più quelli di tipo predittivo e probabilistico come il *machine learning* o il *deep learning*, incrementano la capacità umana di *agency* promuovendo il diritto di autodeterminazione dei soggetti e il potenziamento delle loro possibilità di azione individuali e collettive, dall'altro essi non sono esenti da importanti rischi di natura etica. Studiando in modo particolare il potenziamento algoritmico predittivo basato su tecniche di profilazione delle identità individuali, Tiribelli evidenzia come tali sistemi siano in grado di orientare i comportamenti umani fino a determinare una vera e propria manipolazione ideologica dei singoli.

Rispetto alle numerose ricerche già attive nell'ambito del diritto alla privacy e della tutela dei dati personali nel contesto della cybersfera, la novità della proposta teorica di Tiribelli consiste proprio nel non considerare l'identità personale in termini esclusivamente informativi ma nell'assumere una prospettiva etica che concepisce l'identità come un

processo di maturazione del senso dell'esistere in una prospettiva antropologica interale. Sottolineando a più riprese nel suo volume la necessità di recuperare una riflessione etica sul tema dell'identità personale, Tiribelli propone di riformulare lo stesso diritto alla privacy in termini di diritto alla libertà di scelta e di auto-realizzazione. Secondo la studiosa, infatti, quando si parla di costituzione dell'identità personale si fa riferimento ad una dinamica che è più profonda rispetto alla mera acquisizione di dati ma che riguarda il costituirsi della soggettività nella sua più radicale aspirazione etica. In questo senso, il suo contributo intende offrire un'analisi etica sull'identità personale nel contesto della profilazione algoritmica a partire da una visione antropologica che non concepisce l'essere umano come semplice aggregazione informazionale ma come realtà complessa e storicamente data. Come scrive la studiosa: "il quadro filosofico in cui si muove questo lavoro assume la libertà di scelta e l'autonomia come due elementi fondamentali nel processo di costruzione e sviluppo dell'identità personale, in accordo con una tradizione di ispirazione kantiana, che vede nelle scelte e nelle azioni degli individui non solo il luogo di espressione ma anche il luogo di costituzione e di formazione dell'identità personale" (pp. 14-15). In continuità con la tradizione della filosofia morale di matrice kantiana, Tiribelli fonda la sua analisi sul legame costitutivo che sussiste tra identità personale e libertà di scelta mostrando come una riflessione filosofica di questa natura sia chiamata a confrontarsi sul modo in cui le tecnologie digitali influiscono sull'orizzonte di senso e sul costituirsi delle soggettività personali nella loro apertura assiologica.

Ricco di riferimenti alla letteratura scientifica e alla più recente bibliografia in materia, il volume si presenta articolato in tre capitoli organizzati secondo una logica argomentativa progressiva: nel primo capitolo l'autrice opera un'analisi critica sul tema della libertà personale negli ambienti digitali confrontandosi, in particolare, con quella letteratura che descrive l'identità soggettiva in termini prevalentemente informazionali (Mittelstradt *et. al.*, 2016); qui l'autrice, rifacendosi ai principali paradigmi teorici sul tema della libertà di scelta (Gabbert, Frye, Hartmann, Fabris, De Caro) propone un recupero del concetto etico di libertà come fondamento del processo identitario offrendo un'indagine di quelle condizioni etiche necessarie per l'esercizio della libertà stessa (come ad esempio la disponibilità di opzioni e l'autonomia morale). Nel secondo capitolo, invece, si analizzano in modo più tecnico le modalità con cui le tecnologie digitali algoritmiche incidono sulle condizioni di formazione delle identità personali esercitando un vero e proprio impatto ideologico. In questo senso, Tiribelli esprime la sua preoccupazione etica di fronte ai possibili condizionamenti morali e alle limitazioni delle opzioni di scelta che possono derivare dall'impatto algoritmico sulle vite personali. Tale

impatto, infatti, rischia di tradursi in una de-limitazione dello spazio delle opportunità identitarie dovuto anche ai pregiudizi algoritmici (*bias*) e alla sempre maggiore delega delle scelte etiche che implica una vera e propria erosione dell'autonomia morale. Infine, nel terzo capitolo si offre un bilancio critico della ricerca condotta enucleando la novità dell'approccio etico nella considerazione del tema. Qui Tiribelli, superando i limiti delle teorie informazionali dell'identità, esplicita la caratterizzazione etica del processo identitario mostrando come la libertà morale non possa essere sussunta nelle informazioni o nei dati elaborati dai sistemi algoritmici. Secondo, la studiosa, piuttosto, occorre lavorare ad una progettazione etica di tali dispositivi perché questi siano in grado di promuovere la libertà personale e non di limitarla o addirittura soffocarla. Esulando da ogni considerazione di carattere metafisico sul tema della libertà personale, le riflessioni di Tirabelli nel volume si sbilanciano verso una dimensione pratica attestando l'urgenza di informare di una prospettiva etica la progettazione algoritmica, senza evadere del tutto dal portato delle teorie informazionali sull'identità ma arricchendole di una luce prospettica nuova, quella dell'etica.

Il grande guadagno del volume di Tirabelli consiste allora in una riconsiderazione del tema dell'identità personale nel contesto delle società algoritmiche; tema che non può essere appannaggio esclusivo dell'ambito giuridico e normativo ma deve collocarsi anche in una prospettiva etica. Infatti, come si conclude nel volume, è necessario stabilire criteri etici e azioni specifiche per mitigare le interferenze algoritmiche e per discernere possibilità e rischi di tali sistemi. Solo una progettazione etica di essi, un *ethics by design*, può orientare le società tecnologiche alla promozione della piena identità umana la quale non può essere ridotta ad una aggregazione di dati, ma va colta nella sua complessità bio-psico-spirituale e nel suo legame profondo con la dimensione della libera scelta etica. Superando una concezione meramente informazionale dell'identità a vantaggio di una etica, il volume di Tiribelli si presenta allora come uno strumento teoretico di grande rilevanza per abitare consapevolmente il nostro tempo recuperando le istanze profonde alla base dell'identità umana.

*Finito di stampare
nel mese di ottobre 2024
da Digital Team – Fano (PU)*