DOI: 10.69138/arete-2531-6249-20220714

Responsabilità e cura. Argini, derive e nuove frontiere etiche ai tempi dell'intelligenza artificiale

Aldo Pisano[1]

"Legge zero: Un robot non può recare danno all'umanità, né può permettere che, a causa del suo mancato intervento, l'umanità riceva danno"
I. Asimov, *Io, Robot* (1950)

Sommario: 1. Estendere gli argini. La responsabilità ai tempi dell'IA. 2. Arginare le derive. La responsabilità dell'IA fra affidamento e delega. 3. Esonerare. Sistemi di fiducia e responsabilità. 4. Ampliare le frontiere. L'IA tra micro-sistemi e macro-sistemi. 5. Curare le frontiere. La responsabilità dell'IA nel processo educativo.

Abstract: Aim of this paper is to underline the rule of accountability in global and local challenge referred to Artificial Intelligence. According to an ethical prospective in projecting and applying AI in different fields, we'll focus on systemic and individual moral challenges and their connection. Accountability, according to Hans Jonas' theory has to be extended; and now shared with artificial system that can support humankind in different fields: politics, economics, education. After this first general perspective, we'll focus on the ethical developments of AI in care relationship, assuming Tronto's paradigm of care ethics. Specifically, the paper wants to analyze AI limitations in educational relationships.

1 Laureato in Scienze Filosofiche presso l'Università della Calabria e dottorando in Learning Sciences and Digital Technologies.

Keywords: *Ethics, AI, Accountability, Care, Education.*

1. Estendere gli argini. La responsabilità ai tempi dell'IA

Oggi è lecito chiedersi se Hans Jonas avesse ragione. Dalla pubblicazione de *Il principio responsabilità* è quasi trascorso mezzo secolo. La struttura argomentativa dell'opera apriva e affinava lo spazio del dovere e delle responsabilità umane partendo da quella splendida metafora del "prometeo scatenato"[2]. La tecnica, strutturalmente e trasversalmente oggetto di riflessione del '900, trova oggi un ulteriore estensione del campo di indagine etico-filosofica proprio in riferimento all'intelligenza artificiale.

La relazione fra sapere, potere, tecnica e responsabilità – argomentazione che regge l'intero scheletro del testo di Jonas – attinge da un sostrato ontologico della relazione uomo-natura, tale che essa possa essere facilmente applicata alle attuali emergenze etiche. Le relazioni rese sempre più fluide dalla virtualità, la creazione di realtà parallele (non ultima il Metaverso di Zuckenberg), l'utilizzo indiscriminato delle risorse ambientali e delle macchine ha prodotto effetti che è necessario arginare, fuoriuscendo da quella dicotomia tra principio "speranza" (Bloch)[3] e principio "disperazione" (Anders)[4]. La realtà ormai impone uno svecchiamento da quella presa partitica tra apocalittici e integrati (Eco)[5], per meglio osservare le potenzialità delle nuove tecnologie, definendo e disciplinandone l'utilizzo dal punto di vista etico-giuridico.

In merito all'intelligenza artificiale, a cui oggi si affidano sfide locali e globali, la definizione delle questioni etiche è ancora in corso. Il limite di applicabilità dell'IA è infatti riferibile a una vasta gamma di ambiti: dai veicoli autonomi alle macchine

2 Cfr. H. Jonas, *Il principio responsabilità*, tr. it. di P.P. Portinaro, Einaudi, Torino 2002. Per una trattazione sistematica sul concetto di responsabilità si rinvia all'ottimo testo di C. Bagnoli, *Teoria della responsabilità*, il Mulino, Bologna 2019.

3 Cfr. E. Bloch, *Il principio speranza*, 3 voll., Mimesis, Milano-Udine 2019.

4 Si ricordi che anche Anders ricorreva alla metafora dell'uomo prometeico, sottolineandone la sostanziale obsolescenza rispetto alla macchina. Si vedano: G. Anders, *L'uomo è antiquato*, 2 voll. Bollati Boringhieri, Torino 2002. G. Anders, *Brevi scritti sulla fine dell'uomo*, Asterios, Trieste 2016.

5 Cfr. U. Eco, *Apocalittici e integrati*, Bompiani, Milano 2001.

ecosostenibili, dai robot chirurgici alla robotica educativa. Tutti sistemi a cui può essere affidato un compito e dunque un impegno e una responsabilità dell'azione *nel* e *sul* mondo. A questo proposito tre aspetti diventano oggetto di una prima tematizzazione, per meglio inquadrare il prosieguo del presente lavoro:

- Il concetto di "affidabilità" o "affidamento";
- La correlazione fra "affidamento" e "delega", da cui la necessaria revisione di alcune categorie etiche, quali responsabilità, autonomia, azione, giudizio, scelta e cura;
- L'ambito della cura e la sua relazione con l'IA e in particolare la robotica circoscritta all'ambito educativo.

Per evitare un dilungarsi eccessivo della trattazione, qui non si prenderà in considerazione un problema che è di carattere teoretico, seppure essenziale, che impone di riflettere su cosa si debba intendere per "intelligenza" e su cosa si debba intendere per "artificiale". Il problema di definizione è anche una difficoltà *interna* alla stessa scienza dell'IA e, in generale, alla *data science*. Di fatto, l'origine storica della disciplina "Intelligenza Artificiale"[6] nel celebre convegno di Dartmouth del 1956 apre la strada a una riscoperta del concetto di intelligenza umana, al funzionamento e all'imitabilità dei processi cognitivi – si pensi al diffuso modello HIP (Human Information Processing). Una questione teoretica simile, a confine con gli studi di scienze cognitive[7], rimane un problema abbastanza discusso tra informatici, roboticisti e altre categorie che lavorano tecnicamente sull'IA. Un caso specifico è quello del *deep learning*[8] e la progettazione di modelli di imitazione del

6 Cfr. S. Russell – P. Norvig, *Intelligenza artificiale. Un approccio moderno*, vol. 1, Pearson, Milano 2010. Si vedano anche: J. McCarthy, *What has AI in Common with Philosophy?*, «IJCAI, Proceedings of the 14th International Joint Conference on Artificial Intelligence», vol. 2, agosto 1995; R. Thomason, *Logic and Artificial Intelligence*, in *The Stanford Encyclopedia of Philosophy*, Metaphysics Research Lab, Stanford University, 2020.

7 Si pensi, ad esempio, al problema fra rappresentazione mentale e processo cognitivo. Un interessante analisi del tema è in R. Cordeschi, *Vecchi problemi filosofici per l'intelligenza artificiale*, «SWIF», 1, 23, 2003.

8 Cfr. R. Marmo, *Algoritmi per l'intelligenza artificiale. Progettazione dell'algoritmo, dati e machine learning, neural network, deep learning,* Hoepli, Milano 2020.

linguaggio naturale. Di fatto, qui intervengono temi di matrice wittgensteiniana[9], anche in riferimento al rapporto tra forma di vita e linguaggio. In questi termini, la programmazione del *NLP* (Natural Language Processing)[10] – teso a creare IA in grado di processare il linguaggio in forma naturale – ha messo gli studiosi di IA di fronte ai problemi del simbolico e della semantica tendenzialmente classificando l'intelligenza artificiale come sub-simbolica. L'IA applicata al linguaggio trova ampio spazio nei sistemi di traduzione o nella programmazione di Chatbot sempre più efficienti, in grado di elaborare testi sintatticamente corretti – da questo dipende molto l'implementazione del modello di *NLP*.

Sempre in riferimento a problemi teoretici che sollevano nuove sfide etiche, si pensi all'impatto che avranno i computer quantistici, in grado di analizzare sistemi complessi e con una capacità di simulazione e di predizione molto più accurata, insieme ad altri approcci come il *predictive coding*. Questo a differenza di computer tradizionale che ancora funzionano con leggi classiche, di stampo newtoniano, rendendo l'elaborazione più lenta e superficiale[11].

Questo tipo di riflessione aprirebbe due campi di indagine troppo vasti – uno riferibile alla psicologia cognitiva, alle neuroscienze e alla filosofia de linguaggio[12],

9 Uno dei principi che contraddistingue l'essere umano è il suo essere una specie simbolica, quindi la costruzione e l'interpretazione del sistema simbolico. Da questo il wittgensteiniano "vedere come" (ted. *Sehen als*), inteso come la ricerca di un significato oltre la semplice percezione immediata. Per questo l'IA è catalogata come semanticamente cieca, in quanto incapace di accedere al livello del simbolico. Si rimanda a: L. Wittgenstein, *Philosophische Untersuchungen* (1953), tr. it. di M. Trinchero, *Ricerche filosofiche*, Einaudi, Torino 2009. Si veda anche: D. Sirianni, *Vedere come. Iconismo e pluristabilità tra visione e linguaggio*, Aracne, Roma 2017.

10 Cfr. V. Basile, *From Logic to Language: Natural Language Generation from Logical Forms*, University of Groningen, Groningen, Netherlands 2015.

11 Il rapporto tra fisica newtonia e post-newtoniana, tra modelli macroscopici e nanoscopici apre il divario fra diverse concezioni della realtà stessa. Nella quantistica si è di fronte a una struttura probabilistica e indeterminata del reale, a differenza della necessità che regola i sistemi macroscopici, osservabili a occhio nudo. Per un'analisi del rapporto tra epistemologia e quantistica: G. Cavallini, *La costruzione probabilistica della realtà. Dalla fisica quantistica alla psicologia della conoscenza*, CUEN, Milano 2001; R. Healey, *The philosophy of quantum mechanics: an interactive interpretation*, Cambridge University press, Cambridge 1990.

12 Si rinvia a: R.A. Damasio, *L'errore di Cartesio. Emozione, ragione e cervello umano* (1994), Adelphi, Milano 1995; R.A. Damasio, *Emozioni e coscienza*, Adelphi, Milano 2000; R.A. Damasio, *Neuroscience and ethics: intersections*, «American Journal of Bioethics», 7, pp. 3-7, 2007; N. Levy, *Neuroethics. Challenges for the 21st century*, Cambridge University press, Cambridge 2007; S. Pollo, *La morale della Natura*, Laterza, Roma-Bari 2008; W. Safire, *Visions for a new field of "neurothics"*,

l'altro riferibile all'antropologia filosofica[13] – ma allontanerebbe la riflessione dal suo nucleo etico.

In termini di definizione, dunque, si dovrà fare riferimento a un'idea standard di intelligenza artificiale, per poi trattare in maniera cursoria la tipologia di intelligenza in base alla questione etica proposta. Si rende qui necessario partire dal dato empirico che esiste un insieme più o meno omogeneo di "intelligenze" artificiali, specificamente differenziate. In generale, queste rientrano in quell'idea di IA come mezzo che risolve problemi particolari in base a un determinata programmazione. In sostanza la macchina compie un'azione per raggiungere un determinato obiettivo che, tendenzialmente, è innestato da un operatore umano[14].

Per questo motivo, è possibile distinguere tra *narrow-AI* e *strong-AI*, quindi tra intelligenza artificiale debole e forte. Il passaggio dall'una all'altra determina un aumento della complessità (quindi il numero di variabili poste in gioco)[15], per cui la macchina fornirà risposte/azioni che non fanno parte della programmazione di

---

in S. Marcus (a cura di), *Neuroethics. Mapping the field. Conference Proceedings* (May, 13-14, San Francisco), Dana Press, New York 2002. O. Sacks, *Il fiume della coscienza* (2007), Adelphi, Milano 2018.

13 A questo proposito esistono delle trattazioni imprescindibili ed esaustive: A. Gehlen, *L'uomo. La sua natura e il suo posto nel mondo*, Mimesis, Milano-Udine 2010; A. Gehlen, *L'uomo nell'era della tecnica. Problemi socio-antropologici della civiltà industriale*, Armando, Roma 2003; H. Plessner, *I gradi dell'organico e l'uomo. Introduzione all'antropologia filosofica*, Mimesis, Milano-Udine 2019. Su Plessner si veda: I. Crispini, *Tra corpo e anima. Riflessioni sulla natura umana da Kant a Plessner,* Marsilio, Roma 2004; V. Rasini, *L'eccentrico. Filosofia della natura e antropologia in Helmuth Plessner*, Mimesis, Milano-Udine 2013. Per alcune trattazioni più vicine nel tempo, si consigliano: M. Ceruti, *La danza che crea*, Feltrinelli, Milano 2007; T. Fuchs, *Ecologia del cervello. Fenomenologia e biologia della mente incarnata*, Astrolabio, Roma 2021; Y.N. Harari, *Sapiens. Da animali a dei. Breve storia dell'umanità*, Bompiani, Milano 2017; Y.N. Harari, *Homo Deus. Breve storia del futuro*, Bompiani, Milano 2018; P.H. Lindsay – D.A. Norman, *L'uomo elaboratore di informazioni*, Giunti, Firenze 1987; H.R. Maturana – F.J. Valera, *L'albero della conoscenza,* Garzanti Milano 1999.

14 Stuart Russell, co-autore del manuale per eccellenza sull'intelligenza artificiale insieme a Peter Norvig, nel suo testo *Human Compatible* offre tre definizioni di Intelligenza artificiale determinate dal rapporto macchina-azione-obiettivi: a) «una macchina è intelligente nella misura in cui agendo riesce a ottenere ciò che vuole, in base a ciò che ha percepito» (S. Russell, *Human Compatible. Artificial Intelligence and the Problem of Control*, Penguin, London 2002, pp. 18-19; traduzione mia); b) «Le macchine sono intelligenti nella misura in cui dalle loro azioni ci si aspetta che raggiungano i propri obiettivi» (*ibidem*, p. 41); c) «Le macchine sono utili nella misura in cui ci si può aspettare che le loro azioni raggiungano i *nostri* obiettivi» (*ibidem*, p. 9).

15 Tendenzialmente qui si passa dal livello del Machine Learning a quello del Deep Learning.

base. Esistono, infatti

a. «agenti analitici»[16] che hanno una funzione computazionale semplice: fare calcoli, eseguire micro-azioni programmate che circoscrivono l'azione in termini di prerogative ed effetti della macchina, con conseguente riduzione dell'ambito della responsabilità;

b. «agenti esecutivi» che possono svolgere funzioni più complesse, sempre sulla base di una programmazione, e il cui campo di azione – e dunque responsabilità – si estende: robot chirurgici, droni killer, veicoli autonomi.

c. «*agenti interattivi*» che si riferiscono a IA con protesi simili a quelle di animali o persone (*robot-sitter*), tali che possano avere una funzione sociale più delicata e quindi essere inserite nei contesti di cura.

Quell'estensione della responsabilità prospettata da Jonas pare oggi articolarsi ulteriormente. Non solo si accresce il campo di valutazione delle conseguenze prodotte dall'agire umano, ma anche dell'agire *non-umano*. Dunque, si rende necessario riformulare le categorie di responsabilità e per farlo bisogna valutare i casi concreti e le modalità con cui la responsabilità si dà negli ambiti di applicazione della stessa IA.

Immaginando il campo dell'azione umana e quello affidato alle macchine, si può pensare un sistema di intersezione ed estensione degli ambiti di responsabilità. Se la macchina fosse semi-autonoma, allora la responsabilità dell'essere umano si esplicherebbe in processi di supervisione e controllo continuativi. Tuttavia, l'incidenza etica dell'IA dipende dal fatto che essa si possa rendere *completamente* autonoma rispetto all'attività umana.

Il problema emergente è che un'IA "forte" e autonoma verrebbe investita del ruolo di agente morale, laddove non si formuli un parametro o un criterio, a monte, di gestione delle responsabilità, onde evitare che la macchina si definisca semplicemente amorale e/o si scada in una "cattiva infinità" di attribuzione della responsabilità.

16 P. Dumouchel – L. Damiano, *Vivere con i robot. Saggio sull'empatia artificiale*, Raffaello Cortina, Milano 2019, p. 190.

Se almeno un altro essere umano è coinvolto in quanto oggetto della responsabilità (sia esso nelle generazioni presenti o future), allora qualcuno (o qualcosa) *è chiamato a risponderne*. Essendo la macchina priva di possibilità di rispondere moralmente dell'azione, ne deriva una circoscrizione del concetto di autonomia: «La responsabilità di ciò che fa una macchina autonoma non può essere attribuita alla macchina stessa. L'assenza della coscienza di sé e l'incapacità di agire in base a motivazioni morali sentite come proprie non consentono di considerare come autore di una scelta morale nessuna macchina che possiamo scorgere all'orizzonte delle attuali conoscenze scientifiche e tecnologiche»[17].

Si può ipotizzare un quadro iniziale dell'influenza etica dell'essere umano e dell'IA sul sistema-mondo (gli altri, gli animali e l'ecosistema) in un quadro che passa dal locale al globale e per cui l'etica della prossimità "tradizionale" – rimanendo sulla terminologia di Jonas – tende a estendersi verso sistemi più ampi. Riassumendo, si può affermare che con l'avvento dell'IA, la responsabilità umana raddoppia: non solo deve essere oggetto di valutazione l'agire dell'essere umano, quanto anche quello nei confronti dell'IA, anch'essa ora agente che determina effetti sul mondo. Una delega di questa specie si rende comunque necessaria e non è perfettamente aggirabile, in quanto i problemi "globali" che pongono questioni etiche necessitano di apparati protesici (macchine) che svolgano il lavoro che l'essere umano non è in grado di compiere su larga scala e simultaneamente.

Il limite di applicabilità etica dell'IA potrebbe dipendere proprio da questa differenza fra località e globalità, prossimità ed estensione della responsabilità. Nei micro-sistemi l'autorialità dell'atto può essere circoscritta all'essere umano (rapporto medico-paziente, insegnante-allievo), mentre l'attribuzione di responsabilità su macro-sistemi (es. veicoli autonomia e semi-autonomi, droni killer) necessita di una ridefinizione del paradigma di responsabilità. Si consideri che la questione è già oggetto di revisioni e sistematizzazione giuridica. Sull'utilizzo dell'IA si sono pronunciati nel 2017 due organi nazionali: la CNB e la CNBBSV. Nello specifico, è stato prodotto un lavoro dal titolo *Sviluppi della robotica e della roboetica*. Il VI punto "Raccomandazioni" viene infatti modulato su specifici ambiti: sociale; medico; militare, di polizia e di sorveglianza, giuridico.

Sostanzialmente, questo riporta l'attenzione alla formulazione del *trolley*

17 G. Tamburrini, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Carocci, Roma 2020, p. 137.

problem e quindi al dilemma morale, nella misura in cui nei micro-sistemi etici c'è una maggiore coinvolgimento personale (e quindi emotivo), mentre nei macro-sistemi il coinvolgimento diretto/personale si riduce. Un tema caro a Joan Tronto, soprattutto in *Caring Decmocracy* in cui si misura l'estendibilità del paradigma della cura, considerano le mediazioni istituzionali e l'impersonalità dell'agire.

Quello della Tronto[18] è probabilmente una teoria etica che permette di tracciare il confine di utilizzabilità dell'IA nelle situazioni etiche, nonché di meglio rendere nota la relazione fra affidamento e delega. La cura trova un forte aggancio sull'intelligenza umana, qui intesa come un sistema olistico che mette insieme delle componenti non solo cognitivo-computazionali, ma anche narrative, emotive, somatiche. Questo, infatti, coinvolge altre categorie: coscienza di sé, preferenza, scelta, esercizio critico del giudizio[19]. Dimensioni inesistenti nell'IA che rimane epistemicamente opaca e semanticamente cieca.

2. Arginare le derive. La responsabilità dell'IA fra affidamento e delega

Scopo primo della macchina è quello di esonerare. Cosa ci sia dietro il concetto di esonero è un tema articolato che riporta l'attenzione sull'antropologia filosofica e sulla filosofia della tecnica. Rimanendo, però, nel campo dell'intelligenza artificiale, il termine "robot" – utilizzato per la prima volta da Karel Čapek nell'opera teatrale *I robot universali di Rossum* – è una parola di origine ceca (*robota*) che rimanda al lavoro pesante, forzato[20].

Questa considerazione preliminare permette di analizzare il valore della macchina

18 Cfr. J.C. Tronto, *Moral Boundaries. A Political Argument for and Ethic of Care*, Routledge, New York 1993; J.C. Tronto, *Caring Democracy. Markets, Equality and Justice*, New York University Press, New York-London 2013.

19 Uno degli attuali sviluppi relativi alla possibilità di creare tramite algoritmi delle scale valoriali e dei principi nelle macchine è data dall'idea delle cp-nets (Conditional Preferences – net), attualmente oggetto di studio e teorizzazione. Queste si basano sulla riproduzione delle capacità di scelta e giudizio umani mediante un sistema basato sulle preferenze, mediante la programmazione di un sistema deontologico nella macchina. Cfr. R. Calegiari – A. Loreggia – E. Lorini – F. Rossi – G. Sartor, *Modeling Contrary-to-Duty with CP-nets*, «ResearchGate», 23 Marzo 2020. La ricerca è disponibile al seguente link: https://www.researchgate.net/publication/340134395_Modeling_Contrary-to-Duty_with_CP-nets/link/5eac5f46a6fdcc7050a18084/download].

20 Cfr. K. Čapek – R.U.R. *Rossum's Universal Robots*, tr. it. di A. Catalano, Marsilio, Roma 2015.

nella sua funzione supportiva all'essere umano. Dal momento in cui le macchine, nella loro forma più semplice, sono entrate a far parte del mondo del lavoro si è immediatamente prodotto un effetto di riduzione dei costi che è diventata oggetto di interventi sociali (questo già dalla prima rivoluzione industriale). Capitalismo e tecnicizzazione[21] hanno lavorato di pari passo, con il secondo processo che va sempre più a favore del primo grazie all'introduzione di macchinari utili alla produzione e che ottimizzano quello umano, fino alla completa sostituzione di quest'ultimo.

Questa prima deriva sociale dell'utilizzo dell'IA rappresenta di per sé un campo di indagine che qui non può essere oggetto di trattazione, ma utile a comprendere come la macchina stia diventando sostitutiva dell'attività umana. La regolamentazione etica dell'utilizzo dell'IA, di fatto, ha come effetto positivo quello di un orientamento umano-centrico dell'utilizzo delle intelligenze artificiali[22] (nell'ultimo anno nasce anche in Italia *SIpEIA*, la società italiana per l'etica dell'intelligenza artificiale)[23].

Questo significa porre la riflessione etica e giuridica a tutela e garanzia di un progetto umanistico[24]:

- in senso individuale (micro-sistemico), in cui vengono garantite le sane relazioni di cura nei processi che coinvolgono agenti umani, così da garantire un certo livello di qualità della vita, prediligendo un sistema non-tecnocratico e atto a promuovere lo sviluppo individuale (es. ambito educativo);

- in senso politico-sociale (macro-sistemico) rispettando il valore di una società plurale, un mondo aperto e non regolamentato da decisioni affidate a un algoritmo. Scrive Èric Sadin, argomentando sulle derive assolutistiche dell'IA, di «una voce che si rivolge a noi, [...] dotata di un livello di conoscenza

21 Per una trattazione che assume una curvatura più socio-politica si suggerisce É. Sadin, *Critica della ragione artificiale. Una difesa dell'umanità*, LUISS, Milano 2019.
22 Per l'approccio umano centrico si vedano i seguenti documenti: Fondazione Leonardo, Civiltà delle Macchine, *Statuto etico e giuridico dell'IA*, 2019; *Manifesto di Vienna per l'umanesimo digitale*, Vienna, maggio 2019. In quest'ultimo si parla in maniera esplicita di un Umanesimo Digitale che utilizza un approccio critico perché «influenzi la complessa interazione tra tecnologia e persona, per una società e una vita migliori, nel pieno rispetto dei diritti umani universali» (*ibidem*, p. 2).
23 Il sito è consultabile a questo link: https://sipeia.diag.uniroma1.it/.
24 A questo proposito si rinvia a: P. Donatelli, *La filosofia e la vita etica*, Einaudi, Torino 2020.

> e competenza senza eguali e che ci dispensa continuamente buoni consigli, si vedrà investita di un'autorità e di un'aura tali per cui sarà sempre più difficile, visti i continui perfezionamenti, non percepire questi dialoghi come "naturali" e non prender per "oro colato" qualsiasi sua parola»[25].

L'applicazione dell'IA viene semplificata sulla base di due sistemi (micro e macro), tra i quali si pongono delle relazioni di continuità. L'uno, infatti, influisce sull'altro. Per poter operare a un livello macro-sistemico (che quindi può tutelare e disciplinare quello micro-sistemico) è necessario un sistema di regolamentazione su base nazionale e internazionale che a sua volta definisca i limiti della responsabilità e dell'imputaibilità dell'azione. Non si può pensare di utilizzare dei robot-chirurgici se non esiste una legittimazione giuridica nazionale che ne tuteli l'applicazione sia a favore del chirurgo, sia a favore del paziente. Le forze socio-politiche devono dunque definire i limiti di affidamento all'IA in determinate attività. Recuperando il caso dei robot chirurgici, bisogna considerare che, per quanto possano godere di maggiore precisione, la domanda rimane se questi siano in grado di operare una valutazione che esuli dall'obiettivo impostato in fase di programmazione. Se durante un'operazione chirurgica alla tiroide, il chirurgo – in base alla propria esperienza pratica – si rende conto di una massa tumorale asportabile, decide di agire in base alle possibilità offerte dalla situazione in cui opera, assumendosi la responsabilità, divenendo agente morale e agendo sulla base di una saggezza pratica[26] che è intersezione di conoscenza ed esperienza. Questo permette di tutelare il paziente, in quanto prima la massa tumorale viene asportata, minore sarà il pericolo di diffusione del tumore e creazione di metastasi. Se, tuttavia, la macchina viene programmata per una specifica operazione circoscritta all'esportazione chirurgica della tiroide e quindi non programmata per rendersi conto "ad occhio" (ossia competenza frutto di abilità pratica) della massa tumorale, paradossalmente questa è una contingenza che pone a

25 È Sadin, *Critica della ragione artificiale*, cit., p. 53.

26 Per un'analisi dettagliata della "saggezza pratica" e del deficit fronetico dell'IA: Aristotele, *Etica Nicomachea,* a cura di C. Natali, Laterza, Roma-Bari 2005, p. 231; A proposito si vedano: P. Aubenque, *La prudenza in Aristotele*, Studium, Roma 2018; E. Berti, *Le ragioni di Aristotele*, Laterza, Roma-Bari 1989; C. Natali, *Le virtù particolari nell'"Etica Nicomachea" di Aristotele*, in P. Donatelli – E. Spinelli (a cura di), *Il senso della virtù*, Carocci, Roma 2008; C. Natali, *Aristotele* (2014), Carocci Roma; J. Mcdowell, *Incontinence and Practial Wisdom in Aristotle*, in *The Engaged Intellect. Philosophical Essays*, Harvard University Press, Cambridge, Mass 2009.

rischio il paziente, probabilmente obbligandolo a una seconda operazione e dunque sottoponendolo a stress e a un secondo trauma operatorio.

Come e quando fissare i limiti di affidamento rimane dunque il punto su cui ragionare in termini generali. Questo tenendo conto del limite della macchina di *spiegabilità* dei processi interni fondamentali per l'agire morale:

> Una macchina introspettiva sul piano morale potrà ancora mancare della capacità di aprire un confronto dialogico sul perché di una sua scelta. Gli esseri umani applicano capacità dialogiche di tipo etico quando valutano e dibattono – in *foro interno* o nelle discussioni pubbliche – la bontà di una determinata decisione morale, di una regola che guida il loro comportamento rispetto a un'altra, oppure di una politica etica di più ampio respiro. Ecco un'altra competenza etica per agenti artificiali che non presuppone l'attribuzione dello stato di agente morale responsabile. Acquisendo competenze morali dialogiche, le macchine potranno diventare validi interlocutori degli esseri umani sulle scelte da compiere e sulle politiche etiche da adottare[27].

La garanzia della tutela della qualità della vita e della società democratica dipende strettamente dalla libertà garantita a questi due sistemi. La libertà, a sua volta, è da riferirsi a un "limite di affidamento" delle attività umane all'IA, perché non diventi delega. Da questa affermazione è possibile inferire la differenza che viene qui convenzionalmente utilizzata:

- per affidamento si vuole intendere una mansione di cui l'IA si fa carico ma che tuttavia non implica una totale autonomia e in cui la responsabilità è equamente distribuita fra macchina e operatore umano;

- per delega si intende un'estensione dell'affidamento alla macchina che è assoluto.

Il problema non è quello di non delegare affatto, quanto comprendere e fissare il limite fra i due momenti. È questo non può non passare da una riflessione etico-critica e da una consulenza continua che si pone al centro fra la progettazione, la

27 Tamburrini, *Etica delle macchine*, cit., p. 138. A proposito di azione morale e pensiero come attività dialogica, si rinvia a: H. Arendt, *Responsabilità e giudizio* (2003), a cura di D. Tarizzo, Einaudi, Torino 2010 (2ª ed.).

programmazione e la sua applicazione.

Tendezialmente una scelta macro-sistemica importante che disciplini utilizzo dell'IA dipende dalla possibilità di una consulenza etica a livello aziendale (e non solo). Se l'azione politica nazionale e internazionale si orientasse a garantire risorse economiche e professionali alle aziende, sarebbe possibile pensare uno scenario in cui il "consulente etico" e il "consulente giuridico" si affianchino alle fasi di progettazione prima dell'applicazione dell'IA. La consultazione si pone al centro, in quanto la sua influenza può essere più o meno ridimensionata. Questo a garanzia e tutela della libertà di ricerca scientifica che preceda l'applicazione. La consulenza etica espone il proprio punto di vista influenzando i due momenti:

- in *progettazione/programmazione* è

  (a) sia garante della libertà della scoperta scientifica evitando di inficiarla,

  (b) sia garante della progettazione-programmazione perché si evitino i risvolti anti-etici dell'applicazione dell'IA. La questione è spinosa, perché tendenzialmente tutelare l'uno significa non tutelare l'altro e si rischierebbe solo di utilizzare delle belle formule linguistiche che non hanno risvolti pratici. Un'ipotesi risolutiva potrebbe essere la consulenza etica a monte che direzioni la progettazione in senso morale (per cui a nessuno dovrebbe venire in mente di creare delle IA *ad hoc* che possano nuocere agli esseri umani).

- in *applicazione*

  la questione è meno problematica, in quanto l'influenza è uni-direzionale. Anzi, proprio questo è il punto critico in cui opera la consulenza etica, definendo i limiti e le modalità di applicazione delle macchine secondo principi di benevolenza, non-maleficenza, giustizia, autonomia e spiegabilità[28] – quattro dei quali presi in prestito dalla bioetica.

28 Cfr. L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità e sfide*, Raffaello Cortina, Milano 2022.

A sua volta, e qui il riferimento corre ancora a Jonas, questo dipende dai due insiemi interessati: quello micro e quello macro, a loro volta legati alla necessità di un coinvolgimento diretto o meno dell'agente morale. Il punto è evitare le ottiche della disperazione e dei catastrofismi, ma anche quelli disillusi dell'era idillica dell'IA e della super-intelligenza[29]. Per questo, la prospettiva offerta da Jonas pare estremamente ragionevole da riattualizzare, in quanto va a ridefinire i limiti della responsabilità e rimane un modello teoretico-etico applicabile alla nuova distribuzione delle responsabilità. A questo punto, oltre affidamento e delega, bisogna anche considerare il principio di affidabilità.

Uno dei pericoli meno evidenti è il «pregiudizio (*bias*) positivo di automazione»[30] che induce a guardare alla macchina in quanto infallibile. Dunque, non bisogna porsi a garanzia del criterio di infallibilità, ma quello più realistico della *trustworthiness* (affidabilità) per come sottolineato nella *White Paper – On Artificial Intelligence* firmata dalla Commissione Europea il 19 febbraio 2020.

Nel documento, il richiamo è a una scelta politico-sociale (macro-sistemica) che regolamenti l'utilizzo dell'IA in un'ottica umano-centrica. Un'urgenza già nota e su cui sentenziava Francis Fukuyaman nel 2002 che – a proposito dello sviluppo e delle applicazioni delle biotecnologie – scriveva: «We should use the power of the state to regulate it»[31]. Il riferimento alla *trustworthiness* è a una concezione di fiducia frutto di formazione e informazione critica, garantite da (a) un *risk-based approach*, (b) una prospettiva umano-centrica, (c) rispettosa dei diritti fondamentali dell'individuo.

A questo punto, soggetti alla regolamentazione saranno il programmatore e l'utente. Mentre nel primo caso, le ripercussioni e il livello di intervento si riferisce ad aspetti macro-sistemici (chi programma IA ne prevede l'utilizzo in ambiti diversi della vita, da quelli globali a quelli individuali), nel secondo caso è il singolo utente che richiede tutela individuale in termini di sicurezza nella gestione dei dati, *privacy* e inclusione.

La tutela del diritto individuale è dunque un'esigenza micro-sistemica che richiede interventi macro-sistemici, sempre nell'ottica di un progetto umanistico (e democratico) che miri alla regolamentazione dell'IA. Questo dato è da considerare

29 Cfr. J. Lovelock, *Novacene. L'era dell'iperintelligenza*, Bollati Boringhieri, Torino 2020.
30 G. Tamburrini, *Etica delle macchine*, cit., p. 102.
31 F. Fukuyama, *Our posthuman future. Consequences of the biotechnology revolution*, Farrar, New York 2002, p. 10.

come un'emergenza all'interno delle etiche applicate, modulando gli ambiti di utilizzo dell'IA come quello medico o educativo, focalizzando l'attenzione sul piano qualitativo della vita. Infatti, l'uso di IA operanti in base a meccanismi computazionali, potrebbero rendere sempre più effettivo l'appiattimento del piano qualitativo della vita su quello quantitativo, estromettendo alcuni paradigmi etici (deontologico, etica delle virtù, etica della cura). Per scarto, verrebbe tutto a convergere su un'etica normativa di stampo consequenzialista. Quest'ultima da intendersi nella sua declinazione utilitaristica – un calcolo mezzi-fini, causa-effetto – che trova facilmente corrispondenza nel codice binario secondo cui opera la macchina.

Questo vale nella definizione generale della programmazione e progettazione. Riflettere sul perché si programmi, su come si programmi e quali siano le conseguenze è fondamentale perché si proceda a una progettazione etica dell'IA, ma soprattutto rispettando un criterio fondamentale: la *trasparenza* dei processi che regolamentano la programmazione, per rendere più lineare l'applicazione delle IA. Tuttavia, proprio quando intervengono i casi di delega, per cui l'essere umano ***è*** esonerato dalle proprie azioni, si apre la necessità di distinguere fra le diverse intelligenze artificiali. Un conto, infatti, è delegare agenti analitici in attività di calcolo, statistiche, videosorveglianza etc. etc., altro conto è delegare completamente agenti esecutivi o agenti interattivi, dove il livello di coinvolgimento personale è *eticamente necessario*, per cui risulta difficile pensare che la macchina venga autonomamente investita di queste funzioni. L'IA non è la macchina della prima rivoluzione industriale. Essa è onnipervasiva, nascosta, ormai indispensabile (dall'assistente vocale al navigatore, dal computer alla domotica). Con essa si è avviata una forma di convivenza che richiede un intervento etico di ridistribuzione delle responsabilità.

3. Esonerare. Sistemi di fiducia e responsabilità

Prima di entrare nel merito dell'etica della cura è necessario spendere qualche parola sull'esonero e le questioni sociali ed etiche riferibili alla fiducia nella sua funzione etico-sociale. Torna qui utile la riflessione di Niklas Luhmann che definisce la fiducia come sistema di riduzione della complessità[32]. Le operazioni – dal lavoro

32 Cfr. N. Luhmann, *Fiducia*, il Mulino, Bologna 2002.

meccanico ai sistemi robotici più articolati – vengono *delegate* all'IA, in modo tale che l'uomo possa esserne esonerato. Questo non esclude che ciò con cui il soggetto entra a contatto necessita di essere rivestito di una certa familiarità o, per lo meno, che vi sia un accordo che tuteli l'affidatario, da cui l'importanza del disciplinamento giuridico dell'IA: «Un accordo legale che garantisca maggiore sicurezza a particolari aspettative e che preveda sanzioni, costituisce un fondamento indispensabile per qualsiasi considerazione a lungo termini di questo tipo, riducendo il rischio insito nell'accordare fiducia»[33].

Innanzitutto, il fatto che l'intelligenza artificiale venga prima affidata e poi delegata pone a monte un problema cognitivo riferito alla fiducia che, tendenzialmente, significa rinunciare a una conoscenza in cambio di una sicurezza – nel caso dell'IA garantita dal programmatore. Un problema etico non indifferente: l'individuo, oggi, si trova in una rete onnipervasiva (e anche invisibile) di IA che diventa difficile districare. Nel momento in cui l'utente effettua una ricerca sul web, nel momento in cui scorre dei contenuti sul proprio profilo Facebook, nel momento in cui clicca "mi piace" o acquista on-line un qualsiasi prodotto, un insieme di informazioni vanno a sedimentarsi in un *database* e, in misura più ampia, nel grande bacino dei *Big Data*. Questo sistema di controllo intangibile ***è*** utile a capitalizzare l'informazione estrapolata dall'utente, a reindirizzare i contenuti dei *social* per produrre maggiore dipendenza; è una forza nuova che poggia su solide basi comportamentiste (rinforzo positivo e intermittenza, produzione dopaminica con lo scorrimento della *Home* per aggiornamenti della pagina)[34].

Per arginare i pericoli di questa nuova forza[35] è necessario *informare* gli utenti, in quanto persone, dei processi messi in atto dai siti e dalle app che vengono utilizzate. In qualche modo, si è fatto un passo avanti con la richiesta della pagina web di accettazione delle condizioni, il che, in realtà, è un buon raggiro intellettuale, considerando che l'utente medio che effettua una ricerca su internet, per facilità clicca su "accetta tutto", anziché perdere tempo, sito per sito, nell'accettazione dei *cookies*. L'informazione dell'utente rimane una questione di fiducia. Scrive Luhmann:

33 *Ibidem*, p. 52.
34 Cfr. J. Nida-Rümelin – N. Weidenfel, *Umanesimo digitale. Un'etica per l'epoca delle intelligenze artificiali*, FrancoAngeli, Milano 2019.
35 Cfr. A. Soro, *Democrazia e potere dei dati. Libertà, algoritmi, umanesimo digitale*, Baldini e Castoldi, Milano 2019.

> In linea generale comunque, chi ha fiducia cerca nell'immagine soggettiva che ha del mondo punti di appoggio oggettivi del fatto che la fiducia sia giustificata oppure no. Senza una qualche informazione preliminare è praticamente impossibile avere fiducia. La fiducia implica una sovrabbondanza, uno «scoperto» di informazione, e quindi essa si fonda sul fatto che chi si fida si sappia destreggiare fra aspetti fondamentali, o comunque sia già informato su di essi anche se in modo incompleto e inaffidabile.
> I punti di appoggio su cui si fa leva per formare la fiducia non eliminano il rischio, ma si limitano a ridurlo, in quanto non forniscono alcuna informazione esaustiva del comportamento atteso dalla persona a cui è accordata la fiducia. Essi servono solo come trampolino di lancio nell'incertezza, sia pura limitata e strutturata[36].

Dato che il sistema dell'IA si colloca su tutti i piani di vita compromettendoli (basti pensare all'*Internet of Things*, per cui ogni cosa è informatizzata e dotata di un'intelligenza almeno analitica), bisogna ridefinire il campo di estensione della fiducia. È necessario creare sistemi che rendano informato *veramente* l'utente medio con modalità pratico-operative veloci, campagne di sensibilizzazione che possano rendere le persone coscienti della possibilità di violazione del limite che, per lo più, avviene a un livello inconscio. Gli utenti utilizzano le IA ogni giorno nelle sue varie forme e lo fanno senza sapere che esiste un sistema di spionaggio silenzioso e virulento capace di carpire interessi, informazioni, piaceri; un sistema a cui è possibile condizionare la preferenza di voto (da cui un ulteriore messa in crisi del sistema della democrazia); un sistema che, sulla base di una programmazione errata può generare distorsioni cognitive, errori di *bias* e dunque discriminazione. L'IA, almeno al momento, non è capace di esercitare la sospensione del giudizio, di provare empatia: due requisiti essenziali che svolgono una funzione fondamentale nel raffinamento del senso morale, utile a evitare atteggiamenti discriminatori e che anzi costituiscono l'anticamera a tutti i processi relazionali e di cura. Questo insieme di abilità etiche che nascono dalla combinazione fra conoscenza ed esperienza morale del mondo, che plasmano l'agente morale non sono il prodotto di una mera computazione, ma rinviano a un insieme ben più articolato che rende l'essere umano un sistema olistico irriducibile al codice binario o al modello dell'HIP.

36 *Ibidem*, p. 50.

Se dunque Luhmann considera la fiducia come un fondamento invisibile che facilita lo stare al mondo, i meccanismi di funzionamento delle macchine devono essere sottoposti a revisione prima di ogni affidamento e, ancora di più, prima di ogni delega. Il paradigma luhmaniano va esteso e riconsiderato alla luce del fatto che qui non entrano in gioco enti esterni, pubblici a cui ci si affida in maniera sistemica, ma chiama in causa rapporto inter-soggettivi in cui la stessa IA viene utilizzata. Deve diventare dunque naturale chiedersi cosa succede mentre si inseriscono i dati personali (compresa la carta di credito e l'IBAN) su un sito web. Non si è all'interno di un campo in cui è possibile la "fiducia cieca", dato che i sistemi informatici sono a continuo rischio hackeraggio e l'identità stessa dell'utente è a rischio.

Questo pone un'ulteriore questione etica, relativa al fatto che l'individualità dell'utente viene schiacciata sulla mera somma delle informazioni che si riescono a estrapolare, disconoscendo uno *status* propriamente umano che si regge su una sorta di "olismo qualitativo". Il soggetto si definisce in base alla propria memoria, come esperienza vissuta, all'insieme delle scelte, dei pensieri, del suo modo di essere; è un essere incarnato e storicizzato. Il rischio posto dalla società digitale è una privazione della dignità ontologica, propriamente umana del soggetto partendo da una privazione della sua autenticità, dalla sua de-personalizzazione. Come scrive Letizia Oddo, con l'identità virtuale

> viene rotta la relazione complessa fra il corpo e l'ambiente, soggetto e oggetto divengono enti assoluti, cioè sciolti dalle pratiche e da significati che li costituiscono e caratterizzano [...]. La valenza prospettiva e simbolico che pone in continua relazione fra loro, inconscio e conscio, corpo e mente, viene disarticolata nella dimensione artificiale dell'intelligenza, staccata dalla sua matrice corporea e affettiva, isolata alla sua valenza immaginifica e creativa[37].

La nuova intersezione fra micro e macrosistema deve tenere conto di questa dignità propria del soggetto, in quanto essere storico, culturale o, meglio, *homo symbolicus*. Il sistema della fiducia richiede che vi siano processi di supervisione e controllo, atti a garantire la preservazione della dignità umana, politica e sociale

37 L. Oddo, *L'inconscio fra reale e virtuale. Dopo Jung. Visioni della comunicazione informatica*, Moretti e Vitali, Bergamo 2018, pp. 37-38.

degli utenti in quanto persone. Questa riflessione etico-critica necessita di essere curvata e attuata attraverso le strutture del diritto che garantiscano "trasparenza" e "affidabilità", ossia una tracciabilità e incorruttibilità dei processi che inducono l'IA all'azione. È necessario riscrivere l'insieme delle relazioni umane alla luce di questa importante estensione della tecnica nel sistema delle relazioni, per poi riconsegnarle a un sistema di fiducia, rendendo familiare i meccanismi che regolamentano la progettazione, l'applicazione e la diffusione dell'IA. Parafrasando ancora Luhmann, l'ambiente (come insieme di fenomeni ignoti) si è allargato ed è necessario iniziare a inserirlo in un processo di familiarizzazione che non può essere lasciato *solo* all'informazione e alla conoscenza come attività private, quanto educare e promuovere a livello macro-sistemico (politica, società, istruzione) la conoscenza dell'intelligenza artificiale.

Ora, la questione è se i due criteri di trasparenza e affidabilità, soprattutto il secondo, possano essere rispettati nei processi che determinano esonero totale dell'attività umana, con delega alle macchine. È dunque necessario definire il valore antropologico dell'esonero: «Le peculiari condizioni biologiche dell'uomo rendono necessario sciogliere dal mero presente i rapporti con il mondo [...]. Con dispendio minimo di energie e in prestazioni altissime e liberissime – cioè esonerate – noi siamo capaci di anticiparci e riafferarci, di sintonizzarci e di commutarci, di progettare e pertanto di impegnare la nostra attività *nel lavoro*, in un'azione orientata»[38].

Il punto di forza dell'argomentazione di Gehlen è che il principio di esonero viene esteso non al solo ambito motorio – per cui si presuppone un automatismo – ma anche alla vista, al linguaggio verbale e alla rappresentazione. Per questo, l'autore, successivamente, si riferisce all'esonero come il «*progressivo carattere indiretto* del comportamento umano, il contatto sempre più ridotto, ma anche più sottile, più libero e variabile. Tra l'azione e la sua meta si interpongono articolazioni intermedie le quali a loro volta divengono oggetto di un interesse derivato e più indiretto; fare umano è per noi non già l'uso casuale, per scopi immediati, di uno strumento che capiti sotto gli occhi, sibbene la fabbricazione di uno strumento per uno scopo lontano»[39].

Esonerare significa portare avanti l'evoluzione, partendo dal presupposto che le attività divenute abitudini, ripetitive possono essere delegate ad altri mezzi.

38 A. Gehlen, *L'uomo*, cit., p. 89.
39 *Ibidem*, p. 91.

L'esonero valorizza la "via di mezzo" che qui si tenta di tracciare. Questo significa affermare il valore intrinseco dell'IA come imprescindibile per le attività umane. Aumentando la complessità del sistema-ambiente (cioè i confini dell'ignoto) e aumentando proporzionalmente l'incidenza dell'attività umana su questo stesso sistema, la necessità di controllo e monitoraggio continuativo guarda alle macchine intelligenti come esonero per l'uomo e allo stesso tempo tutela per il mondo.

Si immagini questo dal punto di vista macro-sistemico. Il sistema-ambiente diventa sempre più complesso a causa dell'incidenza umana su quello che si è definito mondo. L'ambiente/mondo è il sistema di complessità massimo in continuo aumento, proprio in virtù dell'azione esercitata da parte dell'uomo su di esso. Una questione apertasi intorno alla seconda metà del secolo scorso in riferimento ai temi di bioetica, etica ambientalista[40] e, in generale, delle etiche applicate[41] come tentativo di ricerca di soluzioni rispetto alle nuove sfide emergenti. Lo stesso Jonas ampliava lo spazio della responsabilità proprio in riferimento alle generazioni future.

Questo ha imposto un'estensione del dovere[42] man mano che l'avanzamento tecnico è diventato sempre più pervasivo della vita umana. Il processo è ormai irreversibile. Se, con la prima rivoluzione industriale – ma volendo retrocedere ancora si può arrivare all'invenzione della ruota – l'uomo tende a esonerarsi da compiti meccanici, rimane tempo e modo di sviluppare altre facoltà e processi: «Tutte le funzioni superiori dell'uomo, in ogni campo della vita intellettuale e morale, ma anche in quello dell'affinamento motorio e operativo, sono pertanto sviluppate grazie al fatto che il costituirsi di stabili e basilari abitudini di fondo *esonera* l'energia in essere originariamente impiegata per le motivazioni, i tentativi, il controllo, *liberandola per prestazioni di specie superiore*»[43]. La parola chiave è dunque apprendimento. Come scrive Stuart Russell: «L'apprendimento conferisce un grande vantaggio evolutivo [...]. L'apprendimento è un bene molto più che per

40 A questo proposito si rinvia a J. Passmore, *La nostra responsabilità per la natura*, Feltrinelli, Milano 1991. La prima edizione del testo di Passmore risale al 1974, cinque anni dopo Jonas avrebbe pubblicato *Il principio responsabilità.*

41 A questo proposito di rinvia a D. Fabris (a cura di), *Etiche applicate,* Carocci, Roma 2018; M. Mori, *Manuale di bioetica. Verso una civiltà biomedica secolarizzata*, Le lettere, Firenze 2013; P. Donatelli, *Il lato ordinario della vita. Filosofia ed esperienza comune*, il Mulino, Bologna 2018.

42 Cfr. L. Fonnesu, *Dovere*, La Nuova Italia, Firenze 1998.

43 A. Gehlen, *L'uomo*, cit., p. 93. Corsivo nel testo.

sopravvivere e prosperare. Esso accelera l'evoluzione»[44].

Seguendo lo sviluppo antropologico nel tempo, bisogna immaginare un processo di ampliamento costante del sistema complessità, sia perché si conosce sempre di più, sia perché la presenza umana modifica l'ambiente in un ricambio organico continuo. Se i primi macchinari della rivoluzione industriale riuscivano ad estendere il braccio dell'uomo su una porzione minima di questa complessità, ora l'intelligenza artificiale copre un'area più vasta. L'attività dell'essere umano, così, riesce a essere sostanzialmente tentacolare rispetto ai vari ambiti che riguardano il micro e il macrosistema.

Rimane però un rischio: il possibile decentramento delle funzioni umane sia come operatori, sia come subenti l'azione dell'IA. E qui si arriva al punto di partenza in riferimento all'argomentazione sulle tipologie di intelligenza artificiale. Esse, nel momento in cui possono auto-apprendere, diventano entità autonome, per cui il rischio è quello di non riuscire a misurare gli ambiti di applicazione. Per questo entra in soccorso la dimensione microsistemica e l'etica della cura come paradigma che regolamenti l'esonero.

4. Ampliare le frontiere. L'IA tra micro-sistemi e macro-sistemi

Nelle dimensioni che si stanno qui considerando, l'attività di supervisione e controllo relativa al mondo e all'azione umana su di esso, dipende da una correlazione che coinvolge sia questioni socio-politiche di più ampio respiro, sia questioni relative alla relazione inter-soggettiva che si sviluppa sul livello delle relazioni personali.

L'azione svolta a livello macro-sistemico coinvolge un maggior numero di persone e passa da una serie di operazioni generali, volte a tutelare il mondo e l'altro in forma più astratta. In questo ambito, le etiche normative tornano utili per la progettazione delle IA. L'aspetto micro-sistemico, che si analizzerà in un momento successivo, invece viene circoscritto alla relazione inter-individuale. Chiaramente le intersezioni sono irrimediabili, soprattutto in riferimento alle misure macrosistemiche e alle sue ripercussioni sulla vita individuale. Si pensi, per esempio, all'adozione delle intelligenze artificiali in ambito educativo: è una scelta

44 S. Russell, *Human Compatible*, cit., pp. 18-19. Traduzione mia.

socio-politica con ripercussioni sulle relazioni inter-soggettive. Lo stesso dicasi per i veicoli e le armi autonome. Le sfide globali richiedono la misurazione delle conseguenze sotto un doppio profilo:

- Limitazione del controllo e del monopolio dell'IA. A questo proposito scrive Antonello Soro: «Forse l'arma più intelligente non è quella sul campo di battaglia, ma quella che "controlla i sistemi di controllo" di una società digitale»[45].

- Linee di programmazione che seguano determinati principi etici.

Inoltre, si possono individuare quattro aree di intervento sempre in senso macro-sistemico:

1. *Sociale.*
   In questa dimensione entra in gioco il principio di rispetto della sicurezza: il lavoro, la privacy, l'utilizzo di mezzi autonomi, la discriminazione (*bias*). Inoltre, assumendo il caso delle ricerche internet, queste hanno un valore *confermativo* delle credenze dell'utente, ulteriormente aggravato dalla carenza di un confronto dialogico sano, dal circolo misinformativo e dall'infodemia. In termini di sicurezza, invece, l'hackeraggio del sistema della *Internet of Things (IoT)*, sarebbe capace di bloccare un'intera Nazione e i risultati in termini di perdita di vite umane sarebbero disastrosi[46]. Parlando, invece, degli errori di *bias*, si pensi al contesto giuridico se l'IA valutasse la personalità su una base scientifico-criminologica e l'attitudine a delinquere. A parte la tendenza lombrosiana retrograda che assumerebbe un processo: «Il rischio insito nel ricorso a simili strumenti di giudizio penale è, infatti, un determinismo penale che faccia degenerare il diritto penale "del fatto" (in base a cui ciò che si giudica è il fatto di reato in sé) in diritto penale per tipo d'autore (secondo il profilo stilato dall'algoritmo), qualora la pericolosità sia desunta solo in basse alla propensione a delinquere espressa

45 A. Soro, *Democrazia e potere dei dati*, cit., p. 33.
46 Cfr. *ibidem*.

da determinati fattori valutati su parametri di regolarità statistica»[47].

2. *Politica*.
In riferimento all'utilizzo di strumenti digitali che evitino il *profiling* e quindi di curvare le scelte dei cittadini come si farebbe a livello commerciale. La virtualità esaurisce lo spazio di discussione e dialogo; questo significa una condanna a morte per la democrazia che nasce dal confronto e ricerca comune sul bene[48]. Inoltre, è qui necessario riflettere su politiche internazionali che limitino il possesso dell'IA. Non da escludere è l'impiego dell'IA in riferimento alle «azioni individuali e collettive da compiere per contrastare il riscaldamento globale, alla luce dei principi dell'etica ambientale e di una ripartizione equa tra gruppi sociali e generazioni dei costi e dei benefici»[49]. La questione etica relativa all'IA manifesta sempre l'intrinseca possibilità di conflitto fra etiche normative. La scelta di quale direzione etica adottare chiama in causa la questione relativa al pluralismo etico. La programmazione si deve far carico non solo delle teorie normative, ma anche del valore che può assumere il caso-esemplare da cui attingere concretamente le informazioni per rendere IA autonoma nelle scelte. Questo rientra maggiormente in un paradigma di etica descrittiva che contribuisce ad aumentare il numero di esempi per la programmazione e per rendere l'IA etica[50].

3. *Economica*.
Qui sempre in riferimento al possesso, alla produzione e alla diffusione dell'IA e della gestione dei dati personali, nonché al condizionamento dell'utente. Il potere dei dati e il loro controllo da parte di determinare società supera ovviamente quello statale, arrivando a ridisegnare i rapporti tra pubblico e privato[51].

4. *Educazione e Istruzione*.

47 *Ibidem*, p. 172.
48 Cfr. *ibidem*, pp. 160-161.
49 G. Tamburrini, *Etica delle macchine*, cit., p. 138.
50 Cfr. *ibidem*, pp. 45-46.
51 Cfr. A. Soro, *Democrazia e potere dei dati*, cit., p. 26.

> Qui l'utilizzo del digitale e dell'IA entra in un territorio di confine molto delicato che deve sia tenere conto delle inclinazioni individuali, sia del valore delle esperienze e delle relazioni nel contesto didattico-educativo, sia dello sviluppo delle competenze – intese come analisi e capacità di giudizio – di cui è manchevole la macchina e che quindi non possono essere oggetto di trasmissione. Il tutto riferito al progetto di vita individuale e della costruzione narrativa del Sé e del mondo.

Questi macrosistemi (politica, economia, società, educazione) sono incardinati fra loro, ma anche con altri microsistemi, in cui si possono individuare determinate relazioni intersoggettive. Il macrosistema "Società", ad esempio, includerà (1) il microsistema relativo al rapporto tra paziente e medico, (2) quello relativo a due individui appartenenti a culture diverse. Ora, se in queste relazioni si pone una mediazione artificiale – per una sorta di effetto allargato del principio di indeterminazione di Heisenberg – la struttura relazionale si modifica ed è inficiata da un'altra presenza che può essere più o meno influente a seconda che si tratti di un'IA debole (assistente vocale) o forte (un robot dalle sembianze umane). Non a caso, Masahiro Mori, per questo secondo caso ipotizzava in un grafico una curva di discesa nota come *uncanny valley*[52]. In questa valle del perturbante, l'IA è percepita in maniera eccessivamente simile all'essere umano, compromettendo la possibile relazione di cura che si intende sperimentare utilizzando l'intelligenza artificiale stessa. Ancora qui ritorna quella relazione tra familiare e non-familiare che la trattazione sulla fiducia di Luhmann poneva come essenziale: «Nella maggior parte dei casi una familiarità profonda impedisce che il problema della fiducia diventi argomento di riflessione. Ma quando una riflessione dovesse avere luogo, sarà proprio la familiarità, nel senso del riporre una cieca fiducia in qualcuno, a venir meno. La mancanza di familiarità si diffonde, persino nei confronti di ciò che ci è più vicino, che viene sospinto dai dubbi in una condizione soverchiante di estraneità»[53]. Il confine tra familiarità come cognizione (o informazione) di alcuni processi sistemici e quello che si sviluppa sul piano della relazione personale diviene egualmente precario quando si mette in dubbio la fiducia accordata, soprattutto nel caso delle IA. Entrano qui in gioco alcuni criteri regolativi per la programmazione

52 M. Mori, *The Uncanny Valley*, «Energy», *7* (4), 1970, pp. 33-35.
53 N. Luhmann, *Fiducia*, cit., p. 51.

dell'IA con specifico riferimento alla teoria del consequenzialismo. La progettazione dell'intelligenza artificiale assume come riferimento la misurazione delle conseguenze in base ai criteri con cui l'IA è programmata. Si prenda in considerazione il caso dei veicoli autonomi. La macchina viene programmata in base a determinati parametri per il riconoscimento di alcuni oggetti presenti sulla strada, in relazione alla presenza dei soggetti a bordo, dei tempi di reazione (frenata e accelerazione) etc. etc. Ora, ammettendo il caso di una collisione inevitabile, l'auto dovrebbe essere in grado di valutare (froneticamente) l'azione migliore da compiersi per nuocere al minor numero possibile di persone, fermo restando che il principio primo è quello di "non nuocere a nessuno" laddove possibile. Dunque, la programmazione della macchina deve considerare già questi primi due principi: il primo riferibile a un paradigma consequenzialista, il secondo a un paradigma etico deontologico (dovere assoluto di non nuocere). Fin qui tutto bene e anzi è anche comodo, considerando che è necessario e sufficiente programmare (eticamente) il veicolo autonomo, negando ogni traiettoria di azione possibile che implichi malevolenza. Ma forse non basta, perché l'estensione dei destinatari del dovere è, come si è visto, molto ampia: la macchina dovrà rispettare l'ambiente e ridurre l'emissione di CO2 (il che è uno dei vantaggi auspicabili, magari con riduzione della produzione dei veicoli privati o grazie alla progettazione di auto elettriche), inoltre dovrà rispettare le possibili dimostranze degli animalisti (è in grado di frenare o decidere di non uccidere un animale su strada laddove possibile?) e via dicendo. Questo, per necessaria conseguenza etica, obbliga alla definizione di doveri assoluti e doveri relativi, con una scala valoriale delle azioni possibili che l'auto può perseguire. Sempre che si possa pensare di programmare la deliberazione in base alle variabili nella situazione etica che richiede immediatezza.

E se si può pensare che alcuni doveri siano solo *prima facie*, quindi apparentemente assoluti, allora ci sarà un'altra condizione etica che potrebbe obbligare l'auto a sacrificare la vita di un soggetto che è a bordo, per evitare di ledere la vita di un gruppo di persone che magari in quel momento sta attraversando la strada. La risposta sarebbe quella di programmare l'auto in modo tale che la macchina eviti di nuocere al guidatore, il che riporta l'attenzione su una questione strutturale e quindi riportando l'interrogativo etico alla progettazione: chi comprerebbe o si metterebbe mai su un veicolo autonomo che potrebbe mettere a rischio la propria vita? In questo senso, è possibile pensare una politica libertaria di più ampio respiro per l'utente (tornando sull'intreccio fra micro e macrosistema). Questa renderebbe

possibile la «scelta di atti supererogatori. Un atto di supererogazione è un atto di generosità morale, non deriva da obblighi morali che ciascuno è ragionevolmente tenuto a rispettare, comporti costi personali straordinari, fino a contemplare la possibilità dell'autosacrificio»[54].

Come si può notare, il quadro proposto come esempio nasce dal numero di variabili poste in gioco, dalla loro valutazione complessiva in funzione dell'obiettivo per cui è progettata la macchina, tenendo conto di rispettare alcuni principi etici, primi fra tutti benevolenza e non maleficenza; il rispetto del principio è a sua volta legato sia alla programmazione, sia al modo in cui la macchina rispetta le norme sociali condivise[55].

Concludendo, l'affidamento o la delega all'IA è una scelta che dipende da alcuni parametri:

- *modalità*, ossia dal tipo di intelligenza artificiale impiegata;
- *quantità* che rende possibile alla macchina un'azione coerente con quanto programmato senza arrecare danni o arrecando il minor numero di danni possibili;
- *tempo*, relativo a quanto intercorre fra il presentarsi di una situazione e la riposta della macchina;
- *qualità*, dipendente dal tipo di risposta che la macchina offre per perseguire un determinato obiettivo o per perseguirlo al meglio o di evitarlo in caso si presenti un dilemma etico;
- *valutazione*, che misura se l'azione della macchina avviene a livello macro-

54 G. Tamburrini, *Etica delle macchine*, cit., p. 43.

55 Ricorda Tamburrini fra le sfide globali: «È stato già ricordato che la disponibilità di veicoli ad autonomia crescente e a guida completamente autonoma potrebbe incoraggiare l'uso dei veicoli privati e maggiori consumi di carburante, in assenza di adeguate politiche industriali e tecnologiche; che questi veicoli avranno bisogno di centri di elaborazione dati – CED o *data center* [...]. Si stima che il dispendio energetico dei CED e delle reti di telecomunicazioni sia responsabile di circa l'80% di emissioni di gas serra attribuibili al settore dell'informazione e della comunicazione, e che in un futuro non lontano fino al 13% delle emissioni antropiche di gas serra potrebbe dipendere da questo settore» (G. Tamburrini, *Etica delle macchine*, cit., pp. 138-139).

sistemico o micro-sistemico.

Rimane sempre il problema di chi sia la responsabilità. Probabilmente non è attribuibile nel caso di un agente esecutivo autonomo. La macchina agisce per supportare nei casi in cui è impossibile l'intervento umano o dove sarebbe assurdo mettere a rischio una terza vita. Fino a che decide di non investire tre persone e di schiantarsi contro un albero, allora è utilizzabile. La collisione inevitabile rimane un dilemma sia per l'uomo che per la macchina, ma almeno nel caso della macchina (se ben programmata) non è messa a rischio l'incolumità psico-fisica del conducente. La supervisione umana rimane un principio necessario, per evitare deleghe nocive.

Per capire il limite della delega bisogna attuare la prova della cattiva infinità della catena di controllo. Se, una macchina, deve essere supervisionata continuamente per evitare di incorrere in errori di bias (es. giudice-robot), e se questa supervisione ha bisogno di un'altra macchina che a sua volta debba essere supervisionata e messa sotto controllo, allora in quel caso si può dire che non è conveniente investire su un IA simile per due motivi:

- errato dispendio di risorse;
- impossibilità di sostituzione di un essere umano nell'attività.

Anche il giudice umano può commettere errori, ma non utilizzando un'IA al suo posto: (a) si è tutelato un posto di lavoro (b) non si sono investite inutilmente delle risorse. Anche questo rientra nella valutazione delle responsabilità. In sostanza avremmo evitato danni macro sistemici (dispendio, lavoro) e anche microsistemici (la vita dell'individuo che viene giudicato).

5. Curare le frontiere. La responsabilità dell'IA nel processo educativo

L'etica della cura viene ora chiamato in causa, soprattutto in riferimento al *caring about*, inteso come mero interessamento che implica l'utilizzo di una mediazione ma che non rende effettivo il paradigma della cura. È il caso dell'assegno in denaro per sostenere i bambini malnutriti in Africa. Questi passano da enti, associazioni, mediatori sociali che, escludendo il contatto diretto, compromettono la possibilità di una relazione di cura autentica. Il coinvolgimento è invece fondamentale per la

Ethics of Care, nel momento in cui comporta fisicità, relazione diretta, impiego di competenze e attenzione da parte del *care-giver*. La mediazione è già estromissione dall'atto della cura, perché non ci si sta prendendo cura (en. *take care of*), ma ci si sta semplicemente interessando a una determinata situazione (en. *care about*).

Ovviamente, questo vale anche per il robot che, se immesso in un contesto relazionale, *simulerà* una condizione empatica, esteriormente ma senza un coinvolgimento intenzionale, mentre è proprio questo che rende etica la relazione di cura: l'incontro fra soggettività, fra *care-giver* e *care-receiver*. L'artificialità di un mezzo non-umano è annullamento della cura, in quanto condivisione.

Uno degli esempi più interessanti che si possono addurre è quello della relazione educativa. Qui, non solo si può trovare traccia della relazione fra i due sistemi, ma si può comprendere come il corretto processo educativo, recuperando la sua origine in riferimento alla *Paideia*, permette di intersecare la formazione individuale ai fini di un corretto inserimento nel contesto sociale, con quello che è definito "progetto di vita" e che fa dell'ambiente educativo una fase di transizione. La stessa Tronto afferma che «The care ethic will have profound effects on political life. It will change our conceptions of citizens and of merit, affect forms of political education, and mobilize some excluded political groups»[56]. La corretta attenzione al processo educativo mira a valorizzare la complessità dell'individuo e a formare la personalità, si inserisce in un meccanismo più ampio che fa delle istituzioni educative (la scuola *in primis*), il microcosmo in cui studentesse e studenti iniziano ad imparare come stare al mondo autonomamente. Svilupperanno personalità, intelligenza, abilità pro-sociali. Per questo il paradigma educativo è quello in cui l'interesse macro-sistemico si intesse con quello micro-sistemico, individuando un *continuum* nella distribuzione della responsabilità educativa che procede dallo Stato al singolo insegnante.

La cura non è dunque un oggetto disciplinare, essa è il modello etico con cui i futuri cittadini imparano a stare al mondo, creando un ambiente socialmente confortevole. Primario, in questo processo, è proprio quello della cura delle emozioni che passa dall'incontro diretto con l'altro che è un incontro fisico, storicizzato, culturale, permettendo l'acquisizione di una competenza comunicativa che passa da sospensione del giudizio, etnocentrismo critico, ascolto attivo, empatia[57].

56 J. Tronto, *Moral Boundaries*, cit., p. 177.

57 A proposito della comunicazione interculturale, si rinvia a L.E. Balboni – F. Caon, *Comunicazione interculturale* (2015), Marsilio, Venezia 2020 (6ª ed.).

Lo scambio fra docente e discente, fra maestro e allievo è una relazione fondamentale per coltivare i cittadini di domani in riferimento all'idea del dovere e della responsabilità estesa, non circoscritta alla prossimità spaziale e temporale. Un processo in cui già Jaeger rintracciava l'essenza dell'educazione greca. L'intreccio fra sfera politica e sfera privata è dunque il labile confine su cui si innesta il processo educativo[58].

L'educatore ha un ruolo altamente performativo non solo per la soggettività, ma per l'individualità in riferimento al suo sviluppo nella vita sociale. Dunque, c'è un rapporto che vede il paradigma educativo generale (macrosistema) intrecciato con l'esigenza individuale di autodeterminazione e autonomia (microsistema). L'educatore si pone al centro di questo processo, di questa transizione fra locale e globale, fra necessità e diritto alla formazione dell'individuo secondo la direzione assunta dal contesto socio-politico-educativo di riferimento e la cura dell'individualità come educazione al pluralismo. Questi due aspetti, collettivo e individuale, sono il principio stesso della società educante democratica: tenere insieme l'uniformazione ai principi e ai valori di una civiltà e lo sviluppo del pensiero critico individuale, massificazione e individualizzazione.

Tutto questo paradigma appena messo in evidenza non si rende possibile senza una relazione di cura diretta che, come si è visto con Tronto[59], necessità di uno scambio empatico-emozionale fra soggettività (docente-discente; discente-discente).

Si ammetta, adesso, che la responsabilità educativa, così stratificata e intersecata, venga affidata, a livello micro-sistemico, all'IA – si entra qui nell'ambito *AIED (Artificial Intelligence in Education)*. In questo paradigma è necessario comprendere i limiti e la possibilità di applicazione dell'IA nell'ambito di riferimento. In una relazione così intesa, la formazione dell'individuo non può essere (1) né oggetto di una mediazione, (2) né considerata alla stregua di una trasmissione nozionistica, sempre considerando l'orientamento dell'educazione alla creazione di competenze.

La domanda che si sono posti Damiano e Dumouchel è se esista un'empatia artificiale e quale può essere il suo scopo, soprattutto prendendo in considerazione

58 Cfr. W. Jaeger, *Paideia. La formazione dell'uomo greco. Alla ricerca del divino* (1933), Vol. 2, La nuova Italia, Firenze 1978, pp. 86-87.

59 A proposito di educazione e democrazia, scrive Tronto: «But public schools are also the ways in which communities sustain themselves. John Dewey's (1993) emphasis on the democratizing effects of education points to the role that schools play in a democratic society. In- deed, education can serve as a buffer against injustice» (J. Tronto, *Caring Democracy*, cit., p. 130).

alcune situazioni come quella della disabilità nel contesto educativo[60]. Come scrivono i due autori:

> Il carattere meramente fittizio delle emozioni dei robot attuali è uno dei problemi a cui l'approccio interno alla modellizzazione robotica dei processi emozionali cerca di trovare soluzione. L'urgenza di risolvere la questione è dovuta alla centralità assunta nell'etica della robotica sociale dall'opposizione tra *emozioni umane genuine* ed *emozioni robotiche simulate*. Si tratta in particolare del dibattito inerente ai progetti "assistivi" della robotica sociale, dedicati alla costruzione dei robot "emozionali" o "empatici" atti al supporto e all'assistenza di individui vulnerabili – in primo luogo bambini con bisogni speciali, anziani in perdita di autonomia e persone ospedalizzate[61].

La domanda relativa alla simulazione delle emozioni, in un contesto in cui una delle competenze è proprio educare alle emozioni, è se la macchina autonoma sia capace di valutare empaticamente la condizione dell'essere umano che ha di fronte. E questo non è un dato che emerge con gli studi attuali sulla psicologia delle emozioni, quanto già con Hume veniva posto in evidenza come una dote naturale, utile alla convivenza con gli altri uomini: «Nessuna qualità della natura umana è più notevole, sia in sé sia nelle sue conseguenze, dell'attitudine che abbiamo a simpatizzare con gli altri, ricevendo attraverso la comunicazione le loro inclinazioni e sentimenti, per quanto differenti o persino contrari ai nostri»[62]. Si consideri bene che Hume evidenzia proprio il ruolo che le la simpatia ha in termini di conseguenze, e che essa è un atto comunicativo implicito: «Odio, risentimento, stima, amore, coraggio, allegria e malinconia: tutte queste passioni le sento più per la comunicazione che per mio naturale temperamento e disposizione»[63].

Il riferimento di Hume è dunque da circoscrivere a una tendenza naturale, la simpatia appunto, che trova poi un'evoluzione più raffinata nel concetto di empatia che è invece frutto del processo educativo e che tuttavia si appoggia a

60 Cfr. P. Dumouchel – L. Damiano, *Vivere con i robot*, cit.
61 *Ibidem*, p. 26.
62 D. Hume, *Trattato sulla natura umana*, tr. it. di P. Guglielmoni, Bompiani, Milano 2016 (4ª ed.), p. 631.
63 *Ibidem*, pp. 631-633.

una tendenza naturale che l'individuo sviluppa in ambiente sociale[64]. Questo decade nel momento in cui entrano in gioco le intelligenze artificiali, di fatto «Il riconoscimento del valore intrinseco di un essere umano (della sua dignità in quanto uomo) sarebbe perciò negato *a priori*, se si affida il suo destino alla discrezionalità di una macchina»[65].

La relazione "Carebot" è dunque asimmetrica, in quanto non comporta uno scambio reale fra soggettività, ma fa del robot lo specchio delle necessità dell'agente umano che vi si interfaccia – magari anche producendo una certa dipendenza. Se il robot è un'agente interattivo inserito nella relazione di cura, la responsabilità della delega della relazione rimane sempre di chi programma. Questi agenti artificiali sono "ingannatori" proprio perché «la loro efficacia si basa sull'ingenuità di chi interagisce con loro – spesso individui fragili ed emotivamente vulnerabili, quali bambini con bisogni speciali o anziani in perdita di autonomia»[66].

Senza negare gli effetti positivi che i robot possono avere nella relazione educativa (es. nel caso dell'autismo), si può sostenere che l'ideale sarebbe affidare ma non delegare all'IA le attività di cura; oppure delegare solo alcuni processi circoscritti alla trasmissione nozionistica. Fermo restando che i bambini coinvolti nella relazione con un robot devono essere consapevoli che l'approccio non è con un altro essere umano o con un'individualità piena: «Se il bambino viene imbrogliato, è certo che chi inganna in questo caso non è la macchina, bensì l'adulto – un altro uomo»[67], al quale si attribuisce la responsabilità dell'atto sostitutivo. Gli interlocutori robotici possono

> solo supportare le relazioni di aiuto e di cura. In questo senso tali agenti robotici sono necessariamente *sostituti* temporanei [affidamento] degli interlocutori umani, la cui azione mira a facilitare la creazione di un legame sociale quando, per una serie di motivi, ci sono significative difficoltà di socializzazione. In simili

64 Oggi, le intuizioni humiane si trovano confermate dallo studio sulle attività dei neuroni specchio. Si vedano: G. Rizzolatti – C. Sinigaglia, *So quel che fai. Il cervello che agisce e i neuroni specchio*, Raffaello Cortina, Milano 2006; G. Rizzolatti, *The mirror neuron system and imitation*, in S. Hurley – N. Chater (a cura di), *Perspectives on imitation from neuroscience to social science*, MIT press, Cambridge (MA), vol. 1, 2005, pp. 55-76.

65 G. Tamburrini, *Etica delle macchine*, cit., p. 93.

66 P. Dumouchel – L. Damiano, *Vivere con i robot*, cit., pp. 108-109.

67 *Ibidem*, p. 159.

> condizioni l'"amore meccanico" e l'empatia artificiale non possono essere concepiti nei termini di emozioni "false", atte a produrre risultati positivi manipolando e ingannando agenti umani vulnerabili. Le emozioni e l'empatia artificiale, nelle interazioni tra robot e umani, generano una dinamica affettiva che conduce a forme di coordinazione socialmente pertinenti[68].

La responsabilità educativa è sì circoscritta all'individuo ma si estende spazialmente e temporalmente alla complessità del processo sociale e l'empatia ne costituisce il fondamento.

68 *Ibidem*, pp. 158-159.