

NEI PROSSIMI NUMERI

Sul volume Etica Teologica della Vita
della Pontificia Accademia per la Vita

Del nuovo Codice Internazionale
di Etica Medica della World Medical Association

Ripensare la definizione di morte?

Previsioni circa la sorte della proposta
di divieto universale di gravidanza per altri

Anno XXXI • 2/2023

BIOETICA

Rivista interdisciplinare

Anno XXXI

n. 2/2023



BIOETICA

Vicolo del Pavone

EDITORIALE – Il divieto universale di GPA tutela
i diritti della donna e del nato?

Mori

STUDI E SAGGI – Il ruolo dei CE nel fine-vita
Sul suicidio medicalmente assistito
Lettera aperta a Enrico Rossi sul perché
è giusto che la sinistra sostenga la GPA

Pantaleone, Sesta, Mori

DISCUSSIONE DI LIBRO –
Figli di un dio minore di Luciano Moia

De Zen, Donatelli, Favi, Moia, Raviolo

INTERVENTI – Resoconto del Convegno di Trieste
16-17 giugno 2023. Comunicati Stampa della
Consulta di Bioetica Onlus

Pitacco

ETICA DELL'IA –
IA e questioni etiche, una mappa dei temi

Raffini e Gorrieri

ETICA E INTEGRITÀ NELLA RICERCA – Il dovere di
contribuire alla ricerca scientifica.
Le frodi e l'impatto sociale della
Research Integrity

Annoni, Grignolio Corsini

€ 20,00

ISBN 978-88-7503-274-6



B



Vicolo del Pavone

BIOETICA

Rivista interdisciplinare

Anno XXXI

n. 2/2023



Registrazione del Tribunale di Piacenza n. 641 del 28 marzo 2007 – Iscrizione al Registro degli Operatori di Comunicazione (ROC) n. 25491 – rivista trimestrale.

ISBN 978-88-7503-275-3

ISSN 1122-2344

© Consulta di Bioetica

e-mail: segreteria@consultadibioetica.org

www.consultadibioetica.org

Casa Editrice Vicolo del Pavone

e-mail: info@vicolodelpavone.it

www.vicolodelpavone.it

Direttore responsabile: Maurizio Mori

Direttore: Maurizio Mori

Condirettore: Demetrio Neri

Comitato di direzione: Carlo A. Defanti, Eugenio Lecaldano, Valerio Pocar.

Comitato scientifico: Guido Alpa (La Sapienza, Università di Roma), Sergio Bartolommei (Università di Pisa), Giuseppe Benaglio (La Sapienza, Università di Roma), Elisabetta Bignamini (Ospedale Regina Margherita), Torino, Guido Boella (Università di Torino), Patrizia Borsellino (Università Bicocca, Milano), Caterina Botti (La Sapienza, Università di Roma), Cinzia Caporale (CNR, Roma), Tiziana Caracri (La Sapienza Università di Roma), Lucia Craxi (Università di Palermo), Giorgio Cosmacini (Università Vita-Salute San Raffaele, Milano), Ines Crispini (Università della Calabria), Norman Daniels (Tufts University), Fabrizio de Francesco (Università di Torino), Lucia De Zen (Irccs Burlo Garofolo, Trieste), Stefano Di Donato (Istituto «C. Besta», Milano), Piergiorgio Donatelli (La Sapienza, Università di Roma), Karina Elmir (Istituto Universitario Italiano di Rosario, Argentina), Luigi Ferrajoli (Università di Camerino), Gilda Ferrando (Università di Genova), Antonino Forabosco (Università di Modena e Reggio Emilia), William K. Fulford (Oxford University), Gianluca Gennarelli (Università di Torino), Giuseppe Giaimo (Università di Palermo), Ranaan Gillon (King's College, London), Alberto Giubilini (Università di Oxford, UK), Mariella Immacolato (Asl1 di Massa e Carrara), John Harris (King's College, London), Helga Kuhse (Monash University), Sergio Livigni (Ospedale san Giovanni, Torino), Michael Lockwood (Oxford University), Sebastiano

Maffettone (Università Luiss, Roma), Tito Magri (Università di Bari), Paolo Martelli (Università di Milano), Alberto Martinelli (Università di Milano), Francesca Minerva (Università di Milano), Alberto Oliva (Universidade Federal do Rio de Janeiro, Brasile), Fabrizio Rufo (La Sapienza Università di Roma), Pietro Rossi (Università di Torino), Giovanni Sartor (Università di Bologna), Peter Singer (Princeton), Marcello Valdini (Società Bio-Giuridica, Piacenza), Marco Vergano (Ospedale San Giovanni, Torino), Daniel Wikler (University of Wisconsin-Madison), Paolo Zatti (Università di Padova).

Segreteria di redazione:

Matteo Cresti, Lavinia Del Corona, Laura Gorrieri, Chiara Mannelli, Marisa Polesana, Daniel Raffini, Roberta Martina Zagarella.

Redazione: Stefano Demuru,

Vicolo del Pavone, Via Petazzi 24,

15053 Castelnuovo Scivria (AL)

e-mail: redazione@vicolodelpavone.it

Amministrazione e abbonamenti:

Consulta di Bioetica onlus

Corso San Maurizio 31/A, 10124 Torino

Tel. +39 393 0451364

e-mail: segreteria@consultadibioetica.org

Abbonamento:

Italia: privati 70 €; istituzioni 85 €

Esteri: privati 90 €; istituzioni 100 €.

Gli abbonamenti decorrono dal gennaio di ciascun anno. Chi si abbona durante l'anno riceve tutta l'annata, compresi gli arretrati. Per i numeri precedenti al 2015 rivolgersi alla Consulta di Bioetica Onlus. Per informazioni sullo stato dell'abbonamento, cambi di indirizzo o eventuali disguidi nella consegna dei numeri della rivista, è attivo un servizio informazioni abbonamenti al numero +39 393 0451364 dal lunedì al venerdì, ore 9-19.

Il pagamento può essere effettuato tramite bonifico bancario IT 48 V 07601 01600 000053902961 o versamento sul c/c n.53902961 intestati a Consulta di Bioetica Onlus.



La pubblicazione è stata realizzata grazie al contributo concesso dalla Direzione generale Educazione, ricerca e istituti culturali del Ministero della Cultura.

INDICE

EDITORIALE

- Il divieto universale di GPA tutela davvero i diritti
della donna e del bambino? 214
Maurizio Mori

STUDI E SAGGI

- Il ruolo dei Comitati Etici sul fine vita dopo la riforma di Gennaio 2023 224
Luca Pantaleone
- Lo scudo e la spada 243
Luciano Sesta
- Lettera aperta a Enrico Rossi (ex gov PD della Toscana)
sul perché è giusto che la sinistra sostenga la GPA solidale 267
Maurizio Mori

DISCUSSIONE DI LIBRO: FIGLI DI UN DIO MINORE

- Percorsi di transizione. Sul libro di Luciano Moia 294
Piergiorgio Donatelli
- Osservazioni sul volume “figli di un dio minore?” Di Luciano Moia 305
José Michel Favi
- Cis-dignità permesse e trans-dignità negate: lo sforzo
di restituire ciò che è dovuto 311
Andrea Raviolo
- Figli minori di un Dio minore? Note sul libro di L. Moia
a partire da un punto di vista pediatrico 319
Lucia De Zen
- Perché nella Chiesa l'accoglienza delle diversità sessuali
rimane terra promessa 325
Luciano Moia

ETICA DELL'INTELLIGENZA ARTIFICIALE

Intelligenza artificiale e questioni etiche	332
<i>Daniel Raffini e Laura Gorrieri</i>	

ETICA E INTEGRITÀ NELLA RICERCA

Il dovere morale di contribuire alla ricerca scientifica: una proposta	352
<i>Marco Annoni</i>	
Le frodi dietro il caso Bendectin e la relazione autismo-vaccini	375
<i>Andrea Grignolio Corsini</i>	

INTERVENTI E DIBATTITI

In UK Charlie, Alfie, ... e qui in Italia tutto ok!?!?	394
<i>Giuliana Pitacco</i>	

COMUNICATI STAMPA DELLA CONSULTA DI BIOETICA

1. Autodeterminazione e sciopero della fame	402
<i>(Mariella Immacolato, Pisa) 6 febbraio 2023</i>	
2. Gestazione per altri è buona e va regolamentata	403
<i>(Maurizio Mori, Presidenza) 20 aprile 2023</i>	
3. Brava Aifa! La contraccezione gratuita è salute	405
<i>(Maurizio Mori, Presidenza) 22 aprile 2023</i>	
4. Perché l'orsa JJ4 non va uccisa	406
<i>(Lavinia Del Corona e Francesca Minerva, Milano) 30 aprile 2023</i>	
5. Sostegno al Pride di Roma	408
<i>(Maurizio Mori, Presidenza) 10 giugno 2023</i>	
6. Bilancio della Sentenza Dobbs un anno dopo	409
<i>(Maurizio Mori, Presidenza) 23 giugno 2023</i>	
7. Sostegno al Pride di Belluno 15 luglio	411
<i>(Davide Mazzon, Belluno) 30 giugno 2023</i>	
8. Perché è importante diffondere le Dat	412
<i>(Silvia Delitala, Cagliari), 12 luglio 2023</i>	
9. Buon 45° compleanno, Louise Brown!	414
<i>(Maurizio Mori, Presidenza), 24 luglio 2023</i>	
10. 45 anni dopo (la nascita di Louise Brown)	416
<i>(Carlo Bulletti, Emilia-Romagna), 24 luglio 2023</i>	
11. Sulla pratica della morte medicalmente assistita: sul 3° caso di suicidio assistito a Treviso	420
<i>(Maurizio Mori, Presidenza), 25 luglio 2023</i>	

ETICA DELL'INTELLIGENZA ARTIFICIALE
a cura di SlpEIA

INTELLIGENZA ARTIFICIALE E QUESTIONI ETICHE

ARTIFICIAL INTELLIGENCE AND ETHICAL ISSUES

Daniel Raffini* e Laura Gorrieri**

1. ALLE ORIGINI DELL'INTELLIGENZA ARTIFICIALE

La nascita del termine intelligenza artificiale (IA) si fa risalire al 1956, in occasione della Conferenza di Dartmouth, alla quale presero parte molti studiosi che in seguito diventeranno centrali nella ricerca sull'IA, come Marvin Minsky e John McCarthy. Lo scopo della conferenza, della durata di sei settimane, era quello di creare un gruppo di lavoro interdisciplinare che potesse costruire un primo modello di intelligenza artificiale, ovvero una macchina in grado di simulare l'intelligenza umana. Il gruppo si propose di lavorare in vari ambiti, come il linguaggio naturale, le capacità di astrazione e la creatività, presupponendo l'idea che ogni aspetto dell'apprendimento e dell'intelligenza potesse essere descritto con precisione e replicato da una macchina¹. Secondo questa visione, l'intelligenza umana poteva essere descritta in maniera soddisfacente con l'aiuto di esperti in diverse discipline, individuando le regole che la governano e convertendole in istruzioni comprensibili da un computer. Il limite principale del progetto risiedeva nel pensare di riuscire a descrivere facilmente e in breve tempo i meccanismi di funzionamento della mente umana, un obiettivo ancora oggi non completa-

* Università di Roma La Sapienza, SIpEIA.

** Università di Torino, SIpEIA.

L'articolo è stato ideato e rivisto nella sua interezza da entrambi gli autori. Gorrieri si è occupata della stesura dei paragrafi 1; 2; 4.1; 4.2; 4.3. Raffini si è occupato della stesura dei paragrafi 3; 4.4; 4.5; 4.6; 5.

1) J. MCCARTHY, M. MINSKY, N. ROCHESTER, C. SHANNON, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence", *AI Magazine*, 27 (4), 1955, p. 2.

mente raggiunto dalle neuroscienze. Gli sviluppi successivi della disciplina, inoltre, hanno dimostrato che i risultati migliori nel campo dei sistemi di intelligenza artificiale non si fondano su sistemi esperti a regole². Nel corso del tempo decade l'idea meccanicistica dell'intelligenza umana ed emerge la natura probabilistica degli algoritmi di intelligenza artificiale.

Nel corso della Conferenza di Dartmouth sono anche stati introdotti i concetti di Machine Learning (ML) e Reti Neurali³; tuttavia, bisognerà aspettare ancora qualche decade per avere dei sistemi di ML soddisfacenti, supportati dal maggior potere computazionale dei computer. Fino a quel momento, il campo dell'intelligenza artificiale conosce delle innovazioni, soprattutto tramite i sistemi esperti, o sistemi a regole, tra i quali il più famoso è sicuramente Deep Blue⁴. Si tratta di algoritmi in cui uno scenario viene descritto in tutte le sue possibili varianti e il software prende delle decisioni in base a questa conoscenza. Il cuore di questo sistema è la collezione di regole, che devono essere esaustive, rappresentando tutti i casi possibili che portano a un certo risultato. I sistemi a regole sono molto costosi da costruire in quanto richiedono il lavoro di un esperto di dominio che descriva i casi che il sistema può incontrare.

I sistemi a regole sono stati al centro della scena dell'intelligenza artificiale per molto tempo, ma al giorno d'oggi sono stati superati dai sistemi di ML. I primi sviluppi in questo campo risalgono alla fine degli anni Cinquanta, con l'invenzione di Perceptron⁵, considerato la prima rete neurale.

2) I sistemi a regole sono sistemi decisionali in cui vengono descritti esplicitamente i possibili casi che il sistema può incontrare. Questi sistemi si fondano sulla struttura "if...then...", ovvero vengono elencate le condizioni (gli *if*) che portano ad una certa decisione (i *then*). Ad esempio, un sistema a regole che classifica le e-mail dividendole tra *spam* e *non spam* potrebbe comprendere delle regole quali "se il mittente non è in rubrica", "se nel messaggio della mail si parla di un premio", "se nella mail ci sono molte parole scritte in maiuscolo" come condizioni per dichiarare una certa mail *spam*. In questi sistemi le regole vengono scritte da esseri umani e il loro scopo è rappresentare tutti i casi possibili che il sistema può incontrare, con le adeguate istruzioni da seguire per ogni caso.

3) J. McCARTHY, art. cit., p. 6.

4) Cfr. <https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/>.

5) Cfr. F. ROSENBLATT, "The perceptron: A probabilistic model for information storage and organization in the brain", *Psychological Review*, 65 (6), 1958, pp. 386–408.

I sistemi di ML apprendono a partire da un numero grandissimo di dati che fungono da esempi e che guidano nella creazione di un modello statistico. Nel ML non troviamo più un sistema di regole puntuali scritte da esseri umani: questo significa anche che la conoscenza del sistema non è più mediata dagli esseri umani, come per i sistemi a regole, ma diventa una rappresentazione che ha senso da un punto di vista statistico. Il ML richiede un grande potere computazionale per poter eseguire molte operazioni in poco tempo. Per questo motivo la sua diffusione è strettamente legata alla capacità computazionale dei processori, la cui crescita lineare è descritta dalla legge di Moore⁶. Con la crescita del potere di calcolo, aumenta anche la potenza delle applicazioni di ML. A partire dal 2010 si è assistito a una impennata nella diffusione degli algoritmi di ML grazie alla disponibilità delle GPU (Graphics Processing Unit), ottimizzate per eseguire calcoli complessi. È in questo momento che nasce il Deep Learning (DL)⁷, la branca del ML e dell'IA che ha più utilizzi ai giorni nostri. Il DL prevede l'uso di particolari tipi di Reti Neurali, che vengono chiamate profonde per la loro struttura a più livelli. Inoltre, se nel ML tradizionale permane una guida fornita dagli sviluppatori, che indicano all'algoritmo le caratteristiche da analizzare tramite la definizione delle *feature*⁸, nel DL l'algoritmo seleziona le caratteristiche in autonomia⁹. Quando gli algoritmi individuano delle *feature*,

6) Cfr. G.E. MOORE, "Cramming More Components onto Integrated Circuits", *Electronics Magazine*, 38, 1965, pp. 114-117.

7) Cfr. J. SEVILLA, L. HEIM, A. HO, M. HOBBAHN, P. VILLALOBOS, "Compute Trends Across Three Eras of machine learning". *2022 International Joint Conference on Neural Networks (IJCNN)*, Padua, Italy, 2022, pp. 1-8.

8) Le *feature* sono le caratteristiche secondo cui viene descritto l'input. Ad esempio, se stessimo cercando di fare un modello che sia in grado di predire prezzi di immobili e ci venisse dato accesso allo storico delle transazioni immobiliari di una certa agenzia, dovremmo selezionare le caratteristiche della vendita che secondo noi hanno una ricaduta sul prezzo finale. Queste potrebbero essere la metratura della casa, l'anno di costruzione, la sua posizione. Mentre invece altre caratteristiche potrebbero essere ininfluenti, come ad esempio il nome di battesimo del venditore o l'orario di stipula del contratto. Per selezionare le *feature* si usano diversi metodi, alcuni dei quali sono automatici, per altri invece si seguono le *feature* scelte per gli algoritmi che in letteratura hanno dato il risultato migliore per un certo task.

9) Quando diciamo che nel DL il modello seleziona le *feature*, intendiamo dire che

queste non sono come quelle che troverebbe un essere umano, ma spesso riguardano delle caratteristiche che per noi sono insensate. Al giorno d'oggi quasi tutte gli strumenti di IA sono applicazioni di ML e le più famose sono Reti Neurali di DL (come, ad esempio, ChatGPT o Dall-e). Tali strumenti richiedono un enorme dispendio di energie perché eseguono moltissimi calcoli, soprattutto nella fase di addestramento in cui devono trovare il valore corretto per miliardi di parametri tramite cicli di tentativi e aggiustamenti.

2. INTELLIGENZA ARTIFICIALE: UNA DEFINIZIONE?

Secondo Minsky, “intelligenza artificiale” è una «suitcase-word»¹⁰, ossia un termine su cui a seconda dei contesti le persone hanno depositato significati diversi. Nonostante non esista una definizione unanime di IA, possiamo partire dalla definizione dell'*Enciclopedia della Scienza e della Tecnica* Treccani, secondo cui l'IA è «la disciplina afferente al più ampio campo dell'informatica che studia i fondamenti teorici, le metodologie e le tecniche che consentono di progettare sistemi hardware e sistemi di programmi software atti a fornire all'elaboratore elettronico prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana»¹¹. Molto spesso l'espressione “intelligenza artificiale” viene usata come sinonimo di ML o di DL, ma si tratta di termini distinti, per quanto imparentati. È vero che l'intelligenza artificiale comprende le applicazioni di ML, che a loro volta comprendono quelle di DL, ma esistono anche algoritmi di IA che non si basano su ML. Dunque, l'IA come disciplina studia una vasta gamma di tecnologie, non sempre simili tra loro; all'interno di un universo

i dati in input vengono inseriti *raw*, non trattati. I modelli di DL agiscono basandosi su funzioni di attivazione, chiamate neuroni, che vengono disposti in file. Con i vari cicli di addestramento del modello, ogni neurone dovrebbe prendersi in carico una *feature*, intesa come una caratteristica la cui presenza lo fa attivare. In questo modo le *feature* emergono autonomamente dal sistema, svincolandosi dal controllo umano e diventando rappresentazioni meramente statistiche.

10) Cfr. M. MINSKY, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, New York, Simon & Schuster, 2007.

11) F. AMIGONI, V. SCHIAFFONATI, M. SOMALVICO, “Intelligenza artificiale”, in *Enciclopedia della Scienza e della Tecnica*, Treccani, Roma, 2008.

tanto variegato, possiamo individuare due caratteristiche che accomunano i sistemi di IA: l'autonomia e l'adattabilità. Un'applicazione di IA è tale se è in grado di prendere decisioni senza l'intervento puntuale di un essere umano e di adattarsi alle richieste dell'utente.

Nei media l'espressione intelligenza artificiale viene usata spesso come sinonimo di termini più specifici o di singole tecnologie. Come prodotti specifici di IA troviamo un panorama piuttosto variegato: si va dalle chatbot, ai robot, le macchine a guida autonoma, i sistemi di riconoscimento facciale, passando per droni killer, suggeritori di contenuti o correttore automatico. Usare la categoria di intelligenza artificiale per indicare un prodotto specifico non è accurato, perché alla base di ognuna di queste applicazioni ci sono tecniche diverse tra di loro: alcuni dei prodotti citati sono relativamente semplici da programmare, altri sono molto sofisticati; alcuni si possono programmare in autonomia con il proprio portatile, altri hanno bisogno di enormi infrastrutture e investimenti¹²; alcuni usano algoritmi esistenti da decenni, altri rappresentano la nuova frontiera della disciplina. Non solo i media fanno un uso non preciso del termine, ma spesso anche le aziende lo utilizzano impropriamente per avere un ritorno di immagine, dunque a fini di marketing.

L'imprecisione che circola intorno all'IA ha anche delle ragioni storiche: quando è stata coniata, l'espressione indicava un campo ancora da esplorare, senza confini e caratteristiche chiare. Da quando si è iniziato a parlare di IA, la disciplina è cambiata molto e continua a farlo in maniera veloce. Il termine cattura una caratteristica sfuggente: la futuribilità, ossia la possibilità che qualcosa di inaspettato si realizzi nel futuro. L'IA si occupa di operazioni che una macchina sembrerebbe non poter eseguire e che poi, però, esegue. Da questo deriva il carattere costante di meraviglia che caratterizza la narrazione contemporanea intorno ai sistemi di IA e che allontana da una contestualizzazione corretta e da una comprensione adeguata. La narrazione intorno alle tecnologie si nutre ancora oggi degli immaginari fantascientifici. Le applicazioni dotate di IA nella fantascienza sono molto spesso robot antropomorfi, così simili nell'aspetto e nelle modalità di interazione agli esseri umani da creare un

12) Ad esempio, per ChatGPT l'azienda OpenAI ha raccolto tramite investitori circa 29 miliardi di dollari, come riportato da Forbes: <https://www.forbes.com/sites/qai/2023/01/27/microsoft-confirms-its-10-billion-investment-into-chatgpt-changing-how-microsoft-competes-with-google-apple-and-other-tech-giants/>.

senso di straniamento. I robot antropomorfi della letteratura usano il linguaggio, hanno un corpo con cui compiono delle azioni, possiedono intelligenza emotiva e sono in grado di apprendere e adattarsi al mondo che li circonda in completa autonomia. Essi rispecchiano un tipo di IA chiamata General AI, o Strong AI, che studia le possibilità di creare un sistema in grado di fare (meglio) tutto ciò che farebbe un essere umano. Si tratta di un approccio che ricalca, in una certa misura, l'ideale proposto alla Conferenza di Dartmouth. Nello sviluppo storico dell'intelligenza artificiale si è imposto tuttavia il modello di Narrow AI, o Weak AI, finalizzato allo sviluppo di software specializzati su task circoscritti: ChatGPT non è in grado di muoversi, Dall-E non può scrivere un paper, l'algoritmo di suggerimenti di YouTube non può comporre una melodia. Va sottolineato che negli ultimi anni sono stati addestrati degli algoritmi di IA multimodali, che possono risolvere problemi afferenti a più di un tipo di input¹³, ad esempio immagine e testo. Sicuramente questo rappresenta un'importante innovazione, ma che si iscrive comunque nel solco della Narrow AI, in quanto il sistema non percepisce il mondo in modo olistico, ma manipola diversi input tramite lo stesso processo, senza integrarli.

3. L'ETICA DELL'INTELLIGENZA ARTIFICIALE

I sistemi di intelligenza artificiale stanno diventando sempre più presenti nella nostra società e nelle nostre vite quotidiane, con un impatto significativo anche su settori centrali come la sicurezza, la salute e l'istruzione¹⁴. Questa pervasività richiede da parte dei soggetti interessati una valutazione etica dell'impatto dei prodotti che vengono rilasciati. Tale valutazione deve essere effettuata a tutti i livelli: da parte di chi sviluppa i sistemi, di chi li commercializza e di chi ne fa uso. Per questo motivo è nata una nuova branca della filosofia, che prende il nome di etica dell'intelligenza artificiale e può essere inserita all'interno del più ampio ambito di quella che a volte viene chiamata

13) Cfr. T. BALTRUSAITIS, C. AHUJA, L. MORENCY, "Multimodal machine learning: A Survey and Taxonomy", *IEEE Transactions on Pattern Analysis and machine learning*, 41 (2), 2019, pp. 423-443.

14) Cfr. P. DONATELLI, "Etica e tecnologie digitali: sfide e svolte", *Bioetica. Rivista interdisciplinare*, a. XXII, n. 1/2019, pp. 35-58.

“tecnoetica”¹⁵. Le peculiarità dei sistemi di IA rispetto alle altre tecnologie esistenti rende necessaria una riflessione ad hoc, che tenga conto della particolarità dell’oggetto di studio. Compito specifico dell’etica dell’intelligenza artificiale è lo studio delle implicazioni culturali, sociali ed economiche derivanti dall’utilizzo dei sistemi di IA. In questo contesto, in Italia è stata istituita nel 2020 SIPEIA (Società italiana per l’etica dell’intelligenza artificiale), con lo scopo di individuare un «coordinamento morale capace di garantire eguaglianza e benessere per i soggetti coinvolti»¹⁶. A livello europeo, nel 2019 sono stati pubblicati gli *Orientamenti etici per un’IA affidabile*, nei quali sono individuati i sette requisiti di base per un’IA etica e affidabile¹⁷:

1. Controllo umano dei sistemi di IA per garantire il rispetto dei diritti umani e il benessere degli utenti;
2. Robustezza e sicurezza degli algoritmi e dei sistemi di controllo;
3. Privacy, controllo e gestione dei dati degli utenti;
4. Trasparenza delle operazioni compiute dai sistemi;
5. Diversità, correttezza, assenza di discriminazione;
6. Benessere sociale e ambientale;
7. Responsabilità.

Ma facciamo un passo indietro, cercando di capire lo statuto dell’etica dell’intelligenza artificiale in relazione all’etica tradizionale. L’etica dell’intelligenza artificiale è una disciplina ancora agli inizi, che sta muovendo i suoi primi passi e non risulta ancora ben definita; inoltre, è un campo in continua evoluzione, perché deve andare di pari passo con gli sviluppi tecnici dei sistemi. Per queste ragioni il suo statuto e la sua indipendenza sono ancora oggetto di dibattito. In un recente articolo, Mori, partendo

15) Per un approccio alla disciplina cfr. L. FLORIDI, *Etica dell’intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina Editore, 2022; F. FOSSA, G. TAMBURRINI, V. SCHIAFFONATI (eds.), *Automi e persone. Introduzione all’etica dell’intelligenza artificiale e della robotica*, Carocci, Roma, 2021; G. TAMBURRINI, *Etica delle macchine. Dilemmi morali per la robotica e l’intelligenza artificiale*, Carocci, Roma, 2020.

16) <https://sipeia.diag.uniroma1.it/>. Nel numero 1/2023 di *Bioetica*, SIPEIA è intervenuta nel dibattito in corso con un Manifesto sui Large Language Model e su ChatGPT.

17) <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. Cfr anche F. GRIGENTI, “La prospettiva etica europea sull’intelligenza artificiale”, *Bioetica. Rivista interdisciplinare*, a. XXII, n. 4/2019, pp. 572-596

dall'avvento della bioetica all'inizio degli anni Settanta come una «nuova etica rispetto a quella tradizionale», propone l'idea che l'etica dell'intelligenza artificiale vada considerata come «un'etica nuovissima rispetto alla nuova etica oggi già ampiamente diffusa» e afferma «l'esigenza di ripensare alcune categorie stesse della moralità a noi nota»¹⁸. Non si tratta semplicemente di applicare l'etica all'ambito dell'intelligenza artificiale, ma di costruire nuovi parametri. Quella dell'intelligenza artificiale potrebbe essere una rivoluzione che – al pari dei grandi cambiamenti storici – determinerà un vero e proprio passaggio epocale. Un passo fondamentale, secondo Mori, è quello di sottoporre a vaglio critico le persistenze culturali del passato che in molti casi ostacolano la piena comprensione e lo sviluppo di ambiti del sapere così innovativi come l'intelligenza artificiale. E lo vediamo, come si è accennato, anche nella narrazione che si fa intorno all'IA: una narrazione che oggi nei media e nel discorso pubblico si divide nettamente tra i disfattisti, che vedono nell'avvento dell'IA la fine dell'umanità, e gli entusiasti, che sopravvalutano le possibilità di tecnologie che ad oggi fondano il loro funzionamento su meccanismi puramente probabilistici. Uno degli scopi primari dell'etica dell'intelligenza artificiale dovrà allora essere quello di vigilare sui meccanismi di interazione tra i nuovi strumenti offerti da queste tecnologie e la società, facendo i conti anche con il grado di accettazione e di rifiuto; soprattutto, dovrà occuparsi di mettere in grado il maggior numero possibile di persone di accedere a una comprensione adeguata di ciò che sta accadendo e dei cambiamenti in atto. Si tratta del primo passo della messa al centro della persona umana che da più parti viene invocata e che rappresenta il filo conduttore dell'etica dell'intelligenza artificiale. Una linea che è stata ben definita nel 2020 in occasione della *Roma Call for IA Ethics*, presso la Pontificia Accademia per la Vita, durante la quale si è stabilita la necessità di «garantire una prospettiva in cui l'IA è sviluppata con uno sguardo non alla tecnologia, ma piuttosto al bene dell'umanità e dell'ambiente»¹⁹. Se intorno allo statuto disciplinare già si pongono molte questioni, che in alcuni casi rimangono aperte, la fase problematizzante (e dunque embrionale) in cui si trova oggi l'etica dell'intelligenza artificiale si rivela ancora

18) M. MORI, «La IAetica come etica nuovissima per un mondo ricreato», in R. GRIMALDI (ed), *La società dei robot*, Mondadori Università, Milano, 2022, p. 302.

19) Ivi, p.310.

più chiaramente se ci accingiamo a entrare più nel dettaglio delle singole questioni. Prendiamo per esempio il problema della responsabilità, uno dei sette punti indicati dall'Unione Europea nel documento del 2019: quando l'IA prende decisioni, chi è responsabile di quelle decisioni? Il proprietario dell'IA, il programmatore o il sistema stesso? Domanda di non facile risposta e che prevede, tra l'altro, un'interazione con la sfera del diritto²⁰.

Se sempre più forte è l'attenzione verso la costruzione di un AI for good, rivolta al bene, questo nasce senz'altro in opposizione al fatto che l'intelligenza artificiale può essere utilizzata per scopi immorali o illegittimi, come la discriminazione di alcuni gruppi di individui rispetto ad altri. Ciò può accadere a causa del *bias* nei dati utilizzati per l'addestramento dell'IA o dei pregiudizi che i programmatori, anche inconsciamente, possono veicolare. Per questo motivo è opportuno garantire che i lavoratori e le lavoratrici impegnati nello sviluppo di questi sistemi rappresentino la totalità della società, in primo luogo dal punto di vista di genere ed etnico²¹. L'accesso libero a sistemi di IA non opportunamente istruiti da un punto di vista etico rende possibile un loro utilizzo malevolo. L'IA potrebbe essere utilizzata, ad esempio, per attacchi informatici, per la manipolazione dell'opinione pubblica, o per la creazione di armi autonome, che agiscono senza supervisione umana. Quello delle tecnologie belliche, d'altronde, è un problema centrale per l'etica della scienza a partire almeno dalla bomba atomica. Tutto ciò rende evidente la necessità che ci siano meccanismi di sicurezza robusti per garantire che l'IA non venga utilizzata per fare il male. L'etica dell'intelligenza artificiale ha il compito, tra gli altri, di ribadire in maniera ferma che le tecnologie di IA devono essere utilizzate solo per scopi pacifici e che si deve pertanto promuovere lo sviluppo di controlli adeguati a prevenire lo sviluppo di armi autonome o altri utilizzi non eticamente sostenibili.

Altro problema rilevante per l'etica dell'intelligenza artificiale è il pericolo di aumentare la disuguaglianza economica e sociale, in un mondo

20) Per una panoramica su diritto e intelligenza artificiale cfr. G. SARTOR, *Intelligenza artificiale e diritto*. Giappichelli, Torino, 2022.

21) Cfr. T. CATARCI, D. RAFFINI, *Dove sono le ragazze dell'ENIAC? Donne e Computer Science: tra cancellazioni e stereotipi*, in G. GIANTURCO, G. BRANCATO (eds.), *Oltre gli stereotipi sulla violenza di genere Approcci, teorie e ricerche*, Sapienza Università Editrice, 2022, pp. 113-124.

già estremamente polarizzato tra pochissimi privilegiati e molti svantaggiati e in cui la fascia di popolazione intermedia tende a peggiorare le proprie condizioni piuttosto che a migliorarle. Sebbene l'idea che l'IA finisca per lasciare le persone senza lavoro sia lontana dalla realtà, l'automazione dei processi potrebbe portare alla perdita di alcuni lavori meno qualificati e in generale a una modificazione nel mondo del lavoro, nella direzione di una concentrazione di ricchezza nelle mani di pochi proprietari di IA²². Bisogna sottolineare che un simile scenario e altri possibili scenari malevoli non sono intrinseci alla tecnologia, ma dipendono unicamente dall'uso che si decide di farne. Il fatto che i sistemi di intelligenza artificiale possano essere più facilmente accessibili per chi detiene già il potere e la ricchezza economica potrebbe determinare un uso che renda ancora più polarizzata la nostra società; per questo motivo la ricerca ha il compito di porre dei paletti etici da implementare direttamente all'interno dei sistemi di IA e i governi e le istituzioni hanno il dovere di regolamentare in maniera più efficace possibile l'utilizzo dell'IA per evitare la creazione di monopoli e l'utilizzo speculativo e garantire che i benefici siano distribuiti equamente.

Considerare l'intelligenza artificiale come uno strumento puramente tecnologico significherebbe ignorare le componenti economiche alla base di questo mercato in forte crescita, che si stima essere di circa \$422 miliardi a livello globale²³. Nonostante le prime innovazioni nel campo siano nate sotto il segno dell'open-source, nel settore della tecnologia si sta assistendo alla crescente privatizzazione dei servizi, come è già successo nel caso di internet²⁴. Anche per l'IA negli anni si è assistito a un accentramento verso poche grandi *big tech*, che hanno attuato una aggressiva campagna di acquisizione della concorrenza²⁵. Questo accentramento, combinato con il valore econo-

22) A questo proposito, OpenAI ha pubblicato uno studio sull'impatto di Chat-GPT sul mondo del lavoro (<https://openai.com/research/gpts-are-gpts>).

23) Cfr. <https://www.bloomberg.com/press-releases/2022-06-27/422-37-billion-global-artificial-intelligence-ai-market-size-likely-to-grow-at-39-4-cagr-during-2022-2028-industry>.

24) Cfr. S. GREENSTEIN, *How the internet became commercial: Innovation, privatization, and the birth of a new network*, Princeton University Press, 2015.

25) United Nations Conference on Trade and Development, *Digital Economy Report 2021, Cross-border data flows and development: For whom the data flow*, New York., 2021, p. 30.

mico del settore, ha spostato un enorme peso politico nelle mani di pochi. Lontani dalle fascinazioni della fantascienza, i veri rischi di oggi riguardano il controllo del mercato del lavoro, la gestione dei dati e l'influenza sulle democrazie. Se per Sadin l'intelligenza artificiale è in grado di «enunciare la verità»²⁶, allora chi detiene il controllo di questi sistemi ha anche il potere di manipolarla o comunque di decidere se e quando divulgarla. Lo sviluppo dell'AI ha poi ripercussioni anche a livello ecologico e umano: come segnala Crawford, «la creazione dei sistemi di IA attuali è strettamente legata allo sfruttamento di risorse energetiche e minerarie del pianeta, di manodopera a basso costo e di dati su amplissima scala»²⁷. Le questioni sul tavolo, insomma, sono molte, e promettono di moltiplicarsi sempre di più. È allora quanto mai necessario affrontare fin da subito le questioni etiche più impellenti, se non altro per tentare di contenere la capacità della ragione economia di annullare ogni spazio di riflessione.

4. ALCUNE QUESTIONI ETICHE

A partire dall'impellenza di cui si è detto sopra, nella presente sezione si espongono senza pretesa di esaustività alcune questioni relative all'etica dell'intelligenza artificiale, come spunto per ulteriori approfondimenti e per fornire esempi di casi concreti di questioni problematiche a livello etico. Come si noterà, in alcuni casi (bias, privacy, explainability) i problemi richiedono specifici interventi nello sviluppo tecnico dei sistemi; mentre in altri le questioni etiche sono di carattere più strettamente sociale o economico, come nel caso dei lavoratori digitali sottopagati, dell'utilizzo dell'IA in ambito medico e dell'umanizzazione degli assistenti virtuali.

4.1. PRIVACY E TRASPARENZA DEI DATI

Una delle principali preoccupazioni etiche intorno all'IA riguarda la privacy e la sicurezza dei dati. L'IA richiede grandi quantità di dati per essere

26) E. SADIN, *Critica della ragione artificiale: Una difesa dell'umanità*, Luiss University Press, Roma., 2019, p. 10.

27) K. CRAWFORD, *Né intelligente, né artificiale: il lato oscuro dell'IA*, Il Mulino, Bologna, 2021, p. 22.

addestrata e ciò può portare a violazioni della privacy, nel momento in cui questi dati sono raccolti in modo non etico o utilizzati con finalità diverse da quelle previste. I dati degli utenti possono essere infatti utilizzati per scopi fraudolenti, come lo spionaggio, il ricatto o il furto d'identità. Per questo motivo è necessario che le informazioni siano adeguatamente protette da eventuali hackeraggi. La trasparenza è un'altra preoccupazione importante in relazione all'IA, dal momento che spesso gli algoritmi di IA producono risultati difficili o impossibili da comprendere e spiegare per gli esseri umani. Ciò rende complicato per gli utenti capire come vengono prese le decisioni e quali siano i fattori che influenzano le scelte e può portare a una mancanza di fiducia nei confronti dell'IA. La questione è indicata con il termine *explainability*, ossia la caratteristica dei programmi informatici di spiegare perché è stata presa una decisione. Questo è specialmente un problema nei sistemi di ML, le cui logiche interne possono apparire oscure per gli esseri umani. Nel caso del DL, ad esempio, l'attivazione di ogni neurone dipende da calcoli puramente statistici, la cui rappresentazione può risultare non sensata secondo logiche umane.

Per affrontare queste preoccupazioni, è importante che le organizzazioni che utilizzano l'IA adottino misure per garantire la privacy e la trasparenza. Ciò può includere l'utilizzo di tecniche di crittografia e di protezione dei dati per proteggere le informazioni personali degli utenti e l'implementazione di processi di audit per garantire che gli algoritmi di IA siano correttamente addestrati e producano risultati affidabili. È necessario poi che gli utenti siano informati sulle pratiche di privacy e trasparenza dell'IA e che siano in grado di controllare l'utilizzo dei dati. Ciò può essere ottenuto attraverso la comunicazione chiara delle politiche di privacy e la possibilità per gli utenti di revocare il consenso per l'utilizzo dei propri dati.

4.2. EXPLAINABILITY E BLACK BOX

La questione dell'*explainability* è centrale per lo sviluppo etico dell'intelligenza artificiale, perché è una delle caratteristiche che rende possibile il primo dei principi di affidabilità IA, cioè il controllo umano. Tutti i sistemi di ML si fondano su l'assunto fondamentale che un certo fenomeno sia approssimabile tramite un modello statistico. Questo modello può essere più o meno raffinato, si va da Naïve Bayes che è modellato sulla probabili-

tà condizionata alle reti neurali profonde che compiono molte operazioni diverse e più complesse. In ogni caso, il motivo per cui un certo modello di ML dà un certo risultato è statistico. Non sempre è possibile risalire a come la decisione è stata presa, soprattutto nel caso del DL in cui l'informazione iniziale passa per molti neuroni²⁸. Quando questa operazione non è possibile si parla di *black-box*, ovvero un modello che fornisce output, ma nel quale non c'è modo di risalire ai percorsi. Per cercare di arginare questo problema, è nata la Explainable AI, o XAI, un campo di studi che si occupa di cercare di dare una spiegazione ai risultati dei modelli di ML:

La ri-emersione di questo campo di ricerca è il risultato diretto dell'irrefrenabile ingresso di IA ed ML in diverse industrie e del suo cruciale impatto nei processi decisionali - senza che sia possibile fornire informazioni dettagliate sul ragionamento che ha portato ad una certa decisione, un certo suggerimento, una certa predizione o una certa azione compiuta dalla IA. Di conseguenza, è presente una pressione sociale, etica e legale che richiede delle nuove tecniche di IA che possano rendere le decisioni che questa prende spiegabili e comprensibili.²⁹

Esistono due approcci per la XAI: il primo prevede di utilizzare degli algoritmi *explainable-by-design*, ovvero per i quali è sempre possibile rintracciare il percorso del dato al loro interno; il secondo è la *explainability-post-hoc*, ovvero il tentativo di rendere più trasparenti gli algoritmi opachi, per i quali non è possibile risalire chiaramente al percorso del dato. Non esiste uno standard universale di *explanation* e diversi studi hanno interpretato il termine in vari modi³⁰. Molti dei metodi utilizzati si basano su una spiegazione puramente

28) Ad esempio, per le reti neurali usate nei Large Language Model – come BERT o GPT-3 – si parla di centinaia di migliaia di neuroni nelle stime più conservative e svariate decine di miliardi in quelle più generose.

29) A. ADADI, M. BERRADA, “Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)”, *IEEE Access*, vol. 6, 2018, p. 52140. Traduzione nostra, on originale: «The re-emergence of this research topic is the direct result of the unstoppable penetration of AI/ML across industries and its crucial impact in critical decision-making processes, without being able to provide detailed information about the chain of reasoning that leads to certain decisions, recommendations, predictions or actions made by it. Therefore, the social, ethical and legal pressure calls for new AI techniques that are capable of making decisions explainable and understandable».

30) Cfr. G. VILONE, L. LONGO, “Explainable artificial intelligence: a systematic

statistica, che non permette ulteriori interrogazioni. Un limite, sottolineato da Adadi et alii, è che a tutt'oggi nel campo della XAI non si tenga abbastanza in considerazione il contributo delle scienze sociali, nonostante le evidenti ricadute psicologiche, cognitive e filosofiche dell'utilizzo dei sistemi di IA³¹.

4.3. BIAS E FAIRNESS

L'*Oxford Dictionary* riporta per la parola bias sia il significato di “pregiudizio”³², che quello più neutro di “difetto in un processo di ricerca”³³. La ricerca scientifica, non solo quella sull'AI, può essere compromessa dai bias quando il risultato è influenzato da fattori sistematici che sono stati involontariamente inseriti durante il processo di selezione o elaborazione dei dati. Quando si parla di bias nell'IA ci si riferisce solitamente a errori che creano risultati discriminatori. I risultati dei modelli dovrebbero quindi essere oggettivamente corretti, ma anche eticamente giusti: questa caratteristica viene definita attraverso il termine *fairness*, ovvero «l'assenza di pregiudizi o favoritismi nei confronti di un individuo o un gruppo, in base a caratteristiche intrinseche o acquisite»³⁴. Il problema dei bias è più delicato quando riguarda categorie socialmente marginalizzate, come nel caso di algoritmi che riproducono pregiudizi presenti nella società. Ad esempio, si è scoperto che gli algoritmi usati nel sistema giudiziario statunitense per stabilire il *risk score* dei detenuti, parametro secondo il quale viene valutata la possibilità di

review”, arXiv preprint arXiv:2006.00093, 2020.

31) A. ADADI, art. cit., p. 52142. Traduzione nostra, in originale: «Rarely in literature, we come across the term “social science” or its derivative [sic], yet explanation is a form of social interaction and clearly, it has psychological, cognitive and philosophical projections [...]. Ideas from social science and human behavior are not sufficiently visible in this field».

32) «A strong feeling in favour of or against one group of people, or one side in an argument, often not based on fair judgement».

33) «The fact that the results of research or an experiment are not accurate because a particular factor has not been considered when collecting the information».

34) N. MEHRABI, F. MORSTATTER, N. SAXENA, K. LERMAN, A. GALSTYAN, (2021), “A survey on bias and fairness in machine learning” *ACM Computing Surveys* (CSUR), 54 (6), 2021, p. 1. Traduzione nostra, in originale: «The absence of any prejudice or favoritism toward an individual or group based on their inherent or acquired characteristics».

ottenere la libertà, discriminavano i detenuti di colore³⁵. Un problema di bias è anche quello di alcuni modelli linguistici e traduttori, che perpetuano gli stereotipi di genere abbinando parole come “doctor” a “man” e “nurse” a “woman”. O ancora, software diagnostici che presentano un problema sistematico di mancata diagnosi nei casi di pazienti con la pelle scura³⁶.

Le cause dei bias possono essere molte: alcuni nascono dai dati in input, che presentano qualche tipo di bias che viene poi trasmesso nella fase di addestramento al modello finale; alcuni nascono dal tipo di algoritmo scelto, che può amplificare certe caratteristiche dai dati; altri ancora derivano dal comportamento dagli utenti che hanno fornito i dati, come nel caso dei *self-selection bias*, in cui il campione di input è rappresentato da volontari che potrebbero non rappresentare la totalità della popolazione³⁷. Capire quando e come si presenta un bias non è un compito semplice, ma è necessario per potersi fidare degli algoritmi di intelligenza artificiale. Per questo motivo negli ultimi anni stanno nascendo dei framework legali per poter garantire la *fairness* dei prodotti di IA, come il regolamento GDPR³⁸ e l’AI ACT dell’Unione Europea³⁹.

4.4. IA IN AMBITO MEDICO

Già si è accennato alla contiguità tra etica dell’intelligenza artificiale e bioetica⁴⁰. L’impatto dei sistemi di intelligenza artificiale influenza le nostre vite e quelle degli altri esseri viventi, come nei casi della gestione di applicazioni mediche o di sistemi per il monitoraggio ecologico. L’intelligenza artificiale

35) Cfr. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

36) Cfr. L. SEYYED-KALANTARI, H. ZHANG, M. MCDERMOTT, “Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations” *Nat Med*, 27, 2021, pp. 2176–2182.

37) Cfr. MEHRABI, art. cit.

38) Cfr. <https://gdpr-info.eu/>.

39) Cfr. <https://artificialintelligenceact.eu/>.

40) Cfr. M. TAI, “The impact of artificial intelligence on human society and bioethics”, *Tzu Chi Med J.*, 32 (4), 2020, pp. 339-343.; J. NABI, “How Bioethics Can Shape Artificial Intelligence and Machine Learning”, *The Hastings Center Report*, v. 48, i. 5, 2018, pp. 10-13.

sta diventando sempre più comune nel settore medico, offrendo nuove possibilità per diagnosi, terapie personalizzate e monitoraggio dei pazienti⁴¹. Per questa ragione il 29 maggio 2020 la Presidenza del Consiglio dei ministri ha reso disponibile il documento *Intelligenza artificiale e medicina: aspetti etici*, stilato dal Comitato Nazionale di Bioetica e dal Comitato Nazionale per la Biosicurezza le Biotecnologie e le Scienze della vita, nel quale si descrive l'utilizzo degli strumenti di intelligenza artificiale in ambito medico e si forniscono delle linee guida⁴².

Uno dei principali vantaggi dell'utilizzo dell'IA in campo medico è la capacità di analizzare grandi quantità di dati e fornire diagnosi più precise e rapide⁴³. Ad esempio, l'IA può aiutare i medici a identificare precocemente le malattie e prevedere la progressione nel tempo. Inoltre, l'IA può aiutare i medici a personalizzare le terapie in base alle caratteristiche individuali del paziente, come la genetica, i fattori di rischio e la storia medica. Ciò può portare a una maggiore efficacia dei trattamenti e a una riduzione degli effetti collaterali. Tuttavia, l'utilizzo dell'IA in campo medico solleva anche alcune preoccupazioni etiche. L'utilizzo di sistemi di IA potrebbe determinare discriminazioni dei pazienti sulla base di fattori come l'età, il sesso, l'etnia o la disabilità; o potrebbe tracciare i dati medici dei pazienti senza il loro consenso, violando la loro privacy. Bisogna tener poi conto anche che i risultati restituiti dai sistemi di IA potrebbero non essere sempre esatti e causare diagnosi o terapie errate. Per questo motivo è importante garantire il controllo umano: l'IA deve essere utilizzata come strumento complementare al lavoro dei medici e non come sostituto della loro esperienza e giudizio.

4.5. SFRUTTAMENTO DEL LAVORO PER I DATASET

Come accennato più volte nel corso di questa trattazione, i sistemi di IA hanno bisogno di grandi dataset di esempi, che devono essere curati da esseri

41) Cfr. C. MANNELLI, "Intelligenza artificiale in ambito sanitario. Riflessioni etiche su explainability e bias", *Bioetica. Rivista interdisciplinare*, a. XXIX, n. 1/2021, pp. 51-60. cfr. anche MANNELLI, *Etica e Intelligenza artificiale. Il caso sanitario*, Donzelli, Roma, 2022.

42) https://bioetica.governo.it/media/4260/p6_r_2020_gm_intelligenza-artificialeit.pdf.

43) Cfr. C. LUVERÀ, "Il machine learning come strumento di supporto nelle decisioni mediche: questioni etiche e prospettive", *Bioetica. Rivista interdisciplinare*, a. XXIX, n. 3/2021, pp. 411-428.

umani e dunque richiedono la mobilitazione di moltissimi lavoratori. Questi lavoratori, che uno studio del 2018 stimava come centomila solo sulla piattaforma Amazon Mechanical Turk⁴⁴, vengono chiamati *crowdworker* e svolgono il lavoro di bassa specializzazione che serve per far funzionare i prodotti di ML:

L'intelligenza artificiale ha bisogno di scienziati ed esperti, ma anche di schiere di operai digitali, che raccolgono il materiale didattico per i sistemi e ne sorvegliano i progressi nell'apprendimento. [...] Lavorano da soli, senza contratto e senza previdenza sociale. Non lavorano a giornata ma a minuto, perché i compiti che le piattaforme d'intermediazione digitale gli affidano spesso si possono svolgere molto rapidamente. Il loro compenso è la somma di contributi da pochi centesimi.⁴⁵

Di recente questa categoria i lavoratori è arrivata all'attenzione dei giornali per il caso di OpenAI, che impiega in Africa personale pagato meno di \$2 all'ora per gestire il funzionamento di ChatGPT⁴⁶. La dipendenza di molti sistemi di intelligenza artificiale da lavoro manuale poco specializzato apre una serie di questioni: in primis si pongono problemi di diritto dei lavoratori, che alcuni sindacati stanno cercando di gestire⁴⁷; ci si può inoltre chiedere se i benefici che l'IA prospetta bilancino i problemi che crea, dovendo ricorrere al contributo di persone che lavorano in condizioni precarie e sottopagate; infine è da considerare che la qualità del lavoro dei *crowdworker*, che sono obbligati a svolgere mansioni ripetitive e ai quali vengono richiesti alti livelli di produttività, è direttamente legata alla qualità finale dei prodotti di ML.

4.6. UMANIZZAZIONE ED EMPATIA

Un'altra questione centrale e spesso trascurata che l'etica dell'intelligenza artificiale deve affrontare è quello dell'umanizzazione degli assistenti artificiali che vengono installati nei dispositivi personali o di domotica. Questi spesso

44) D. DIFALLAH, E. FILATOVA, P. IPEIROTIS, "Demographics and Dynamics of Mechanical Turk Workers". *WSDM '18: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 2018, pp. 135.

45) S. BETSCHON, "I precari dei robot", *Internazionale*, 1329, p. 66.

46) Cfr. F. NGILA, "OpenAI underpaid 200 Kenyans to perfect ChatGPT – then sacked them", *Quartz*, 2023. <https://qz.com/open-ai-underpaid-200-kenyans-to-perfect-chatgpt-1850005025>

47) Cfr. <https://faircrowd.work/unions-for-crowdworkers/>.

dotati di una voce umana e in alcuni casi persino di una personalità. Quando parliamo con Alexa o con Siri abbiamo la sensazione di parlare con delle persone, perché si tratta di strumenti che utilizzano una voce e sono studiati per simulare il linguaggio degli esseri umani. Tuttavia, si tratta di sistemi che non hanno l'empatia, la sensibilità e la comprensione che avrebbe un essere umano. Gli utenti, quindi, potrebbero sentirsi vicini agli assistenti virtuali e in alcuni casi stabilire con essi una connessione di tipo empatico. Tale evenienza risulta problematica specialmente in casi di utenti con difficoltà psicologiche o cognitive. Cosa succede se un utente espone a un assistente vocale un disagio psicologico? O se finisce per utilizzare questi strumenti come sostituti della vera socializzazione, magari in situazione di solitudine? Nella maggior parte dei casi gli assistenti virtuali non sono pronti a tali eventualità e nella migliore delle ipotesi danno consigli molto generali e spersonalizzati, come quello di consultare con un esperto. Il problema si pone anche con sistemi che interagiscono in forma scritta. Quando ChatGPT scrive esprimendo dispiacere o scusandosi per un errore sta simulando un'empatia e una comprensione che non esistono. Questi sistemi, inoltre, presentano delle barriere di fronte ai rischi più comuni, ma non riescono a riconoscere in maniera efficace un rischio non calcolato, perfino se è l'utente stesso a dichiararlo. L'etica dell'intelligenza artificiale ha il compito anche di valutare i rischi dell'umanizzazione dei sistemi di IA e in generale di garantire una risposta efficace in casi in cui gli utenti esternino un disagio o un pericolo.

5. CONCLUSIONI

Ripercorrere la storia della disciplina e cercare di decifrare gli impatti sul presente rende evidente come quello dell'intelligenza artificiale sia un campo applicativo con evidenti ricadute etiche. Il presupposto stesso dell'intelligenza artificiale, d'altronde, tocca corde che confinano con la metafisica, sfiorando il desiderio dell'uomo di farsi creatore. Il percorso difficile, fatto di fallimenti e ricalibrature, che porta dall'ideale utopico della Conferenza di Dartmouth fino alle tecnologie che oggi abbiamo a disposizione, è d'altronde la concretizzazione stessa di un rovello che iscrivendosi nelle coscienze degli uomini assume necessariamente una dimensione etica. La via d'uscita sta

nella collaborazione tra diverse aree del sapere, nel dialogo tra l'ambito tecnologico e l'ambito umanistico, nel contributo determinante che possono portare gli studi filosofici e sociologici, come guida per uno sviluppo sostenibile dell'intelligenza artificiale. Questo confronto è reso quando mai necessario dalla posta in gioco. Secondo Crawford i sistemi di IA stanno definendo oggi un «processo di rimappatura e intervento nel mondo»⁴⁸, un processo, dunque, politico. In quanto processo politico, la ridefinizione del mondo messa in moto dall'intelligenza artificiale deve essere resa comprensibile a tutti e deve porsi al centro di un dibattito pubblico consapevole: se ciò non accadrà, il rischio è che diventi un ulteriore strumento di oppressione. Ridefinire l'etica secondo i parametri dell'etica dell'intelligenza artificiale è oggi uno dei compiti più urgenti del nostro tempo. L'IA ha ricadute su tutti noi e questo ci chiama ad uno sforzo comunitario di analisi, sensibilizzazione e studio.

ABSTRACT

La prima parte ripercorre la storia dell'intelligenza artificiale come disciplina e cerca di fornire una definizione. La seconda esamina alcuni problemi e difficoltà che si presentano nel settore. Nell'ultima parte si considerano alcune questioni etiche specifiche come la tutela della privacy, i bias, l'explainability, la fairness, i crowdworker, l'umanizzazione degli assistenti virtuali e alcuni problemi posti dalle applicazioni in ambito medico.

The first part is about the history of artificial intelligence as a discipline and attempts to provide a definition. The second one examines some problems and difficulties of the field. In the last part some specific ethical issues are considered such as privacy, bias, explainability, fairness, crowdworker, humanization of virtual assistants, and some problems concerning applications in the medical field.

KEYWORDS

Intelligenza artificiale • SlpEIA • Privacy
Explainability • Crowdworker

Artificial Intelligence • SlpEIA • Privacy
Explainability • Crowdworker

48) K. CRAWFORD, op. cit., pp. 26-27.