



Intelligenza artificiale e diritto

Come regolare un mondo nuovo

a cura di Antonio D'Aloia



FrancoAngeli

Informazioni per il lettore

Questo file PDF è una versione gratuita di sole 20 pagine ed è leggibile con



La versione completa dell'e-book (a pagamento) è leggibile con Adobe Digital Editions. Per tutte le informazioni sulle condizioni dei nostri e-book (con quali dispositivi leggerli e quali funzioni sono consentite) consulta [cliccando qui](#) le nostre F.A.Q.



BIOETICA SCIENZA SOCIETÀ



Collana fondata e diretta da Antonio D'Aloia

Nata nell'ambito del Centro di Bioetica dell'Università di Parma, la Collana si pone come luogo di riflessione e discussione sulla pluralità di temi riconducibili oggi alla bioetica, dai progressi della scienza e della tecnologia medica alle questioni della sopravvivenza e della sostenibilità ambientale, secondo un approccio interdisciplinare e in una prospettiva di integrazione delle competenze.

In linea con le più aggiornate esperienze internazionali di ricerca e formazione specialistica, la Collana affronta le complessità di questi nuovi sviluppi, i dilemmi etici e le inquietudini sociali che essi suscitano, le sfide che rappresentano per la regolazione giuridica.

Comitato scientifico

Roberto Andorno; Vincenzo Baldini; Roger Brownsword; Stefano Canestrari; Marta Cartabia; Carlos María Romeo Casabona; Carlo Casonato; Alfonso Celotto; Beatrice Centi; Lorenzo Chieffi; Roberto Giovanni Conti; Marilisa D'Amico; Mario De Caro; Roberto Baptista Dias da Silva; Carla Faralli; Giovanni Maria Flick; Tommaso Edoardo Frosini; Silvio Garattini; Michele Guerra; Stefano Giaime Guizzi; Juan Alberto Lecaros Urzúa; Luigi Naldini; Gianluigi Palombella; Andrea Patroni Griffi; Annamaria Poggi; Ulderico Pomarici; Carlo Alberto Redi; Fernando Rey Martínez; Antonio Ruggeri; Carmela Salazar; Amedeo Santosuosso; Andrea Simoncini; Paula Siverino Bavio; Mariachiara Tallacchini; Massimo Villone; Lorenza Violini; Laura Westra.

I volumi pubblicati nella Collana sono sottoposti a un processo di peer review che ne assicura la validità scientifica.

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio “Informatemi” per ricevere via e-mail le segnalazioni delle novità.

Intelligenza artificiale e diritto

Come regolare un mondo nuovo

a cura di Antonio D'Aloia

Scritti di

D. Amoroso, L. Aulino, C. Casonato, A. Celotto, A. D'Aloia,
A. De Felice, L. Dei Cas, F. Donati, F. Faini, M. Fasan, E.A. Ferioli,
R. Folgieri, E. Gragnoli, P. Grimaldi, M.F. Monterossi, F. Pacini,
S. Quattrocolo, L. Rufo, F. Sarzana di Sant'Ippolito, A. Simoncini,
L. Solum, G. Tamburrini, M. Zanichelli.



FrancoAngeli

Il presente volume raccoglie gli atti del Seminario su “*Intelligenza artificiale e diritto. Come regolare un mondo nuovo*”, che si è svolto a Parma il 12 ottobre 2018. Ai lavori presentati in quella sede, si sono aggiunti altri saggi, che hanno consentito di arricchire il quadro dei profili affrontati. Il Seminario è stato voluto dall’Associazione Italiana dei Costituzionalisti, come primo Seminario nell’ambito del Progetto dedicato al tema “*Il costituzionalismo e le sfide del futuro*”, ed è stato organizzato in partnership con lo University Center for Bioethics (UCB) dell’Università degli Studi di Parma, e con il patrocinio del Centro Studi Affari Europei e Internazionali (CSEIA) del medesimo Ateneo.

Il volume viene pubblicato grazie a un finanziamento della Fondazione Cariparma.

Il curatore ringrazia tutti questi Enti per il contributo ideale e materiale che ha reso possibile lo svolgimento del Seminario e la pubblicazione del volume, e in particolare i professori Massimo Luciani, Ginevra Cerroni Feroni, Anna Maria Nico, Lorenza Violini, Giorgio Grasso, Fulvio Pastore, Emanuele Rossi, Antonio Saitta, componenti – insieme al curatore – del Comitato Scientifico del Seminario; e la dott.ssa Maria Chiara Errigo per il coordinamento organizzativo del Seminario e della pubblicazione.

In copertina: Umberto Boccioni, Visioni simultanee
(Von der Heydt-Museum Wuppertal; foto: Medienzentrum Wuppertal, Antje Zeis-Loi)

Copyright © 2020 by FrancoAngeli s.r.l., Milano, Italy.

L’opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d’autore. L’Utente nel momento in cui effettua il download dell’opera accetta tutte le condizioni della licenza d’uso dell’opera previste e comunicate sul sito www.francoangeli.it.

INDICE

<i>Il diritto verso “il mondo nuovo”. Le sfide dell’Intelligenza Artificiale</i> , Antonio D’Aloia	pag. 7
<i>Ecosistemi, opacità, autonomia: le sfide dell’intelligenza artificiale in alcune proposte recenti della Commissione Europea</i> , Maria Zanichelli	» 67
<i>La questione del controllo umano significativo sui sistemi robotici ad autonomia crescente</i> , Daniele Amoroso, Guglielmo Tamburrini	» 89
<i>Artificially Intelligent Law</i> , Lawrence B. Solum	» 117
<i>Per una intelligenza artificiale costituzionalmente orientata</i> , Carlo Casonato	» 131
<i>L’algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà</i> , Andrea Simoncini	» 167
<i>I robot possono avere diritti?</i> , Alfonso Celotto	» 205
<i>La legislazione internazionale in tema di intelligenza artificiale: dai diritti civili alla soggettività della persona elettronica</i> , Fulvio Sarzana di S. Ippolito	» 217
<i>Intelligenza artificiale e giustizia</i> , Filippo Donati	» 237

<i>Equo processo penale e sfide della società algoritmica</i> , Serena Quattrocolo	pag. 267
<i>Intelligenza artificiale e giustizia: tra nuove soggettività giuridiche e problematiche etiche e deontologiche</i> , Livia Aulino	» 283
<i>Stalking e bullismo nell'era dei social network e strumenti di prevenzione dei fenomeni attraverso l'intelligenza artificiale</i> , Paola Grimaldi	» 297
<i>Legaltech, Neurolaw and regulations on artificial intelligent applications: a critical review</i> , Raffaella Folgieri, Ludovico Dei Cas	» 319
<i>Intelligenza artificiale e pluralismo: uso delle tecniche di profilazione nello spazio pubblico democratico</i> , Marta Fasan	» 345
<i>Intelligenza artificiale e decisione politica: qualche considerazione tra questioni vecchie e nuove</i> , Fabio Pacini	» 367
<i>Intelligenza artificiale, diritto e pubblica amministrazione</i> , Fernanda Faini	» 385
<i>Intelligenza artificiale e processi decisionali automatizzati: GDPR ed ethics by design come avamposto per la tutela dei diritti umani</i> , Alfredo De Felice	» 415
<i>Digitalizzazione, Intelligenza artificiale e Robot nella tutela della salute</i> , Elena A. Ferioli	» 423
<i>L'intelligenza artificiale in sanità: tra prospettive e nuovi diritti</i> , Luigi Rufo	» 451
<i>Il rapporto di lavoro, le decisioni automatiche e i loro limiti. La regolazione di fenomeni incomprensibili</i> , Enrico Gragnoli	» 461
<i>Intelligenza artificiale e blockchain: implicazioni reciproche</i> , Maria Francesca Monterossi	» 475

IL DIRITTO VERSO “IL MONDO NUOVO”. LE SFIDE DELL’INTELLIGENZA ARTIFICIALE*

*Antonio D’Aloia***

SOMMARIO: 1. Tra fiction e diritto. Alla ricerca di principi e regole per l’AI. - 2. AI, Robotics, mondo “digitale”. - 3. Perché il diritto deve occuparsi di AI. Dalla cibernetica giuridica al diritto per i sistemi che imparano da soli. - 4. Segue: Macchine (sempre più) autonome e responsabilità. - 5. AI, diritto, diritti. Opportunità e problemi. - 6. Algoritmi e (rischi di) discriminazioni. - 7. AI nel processo e nel procedimento amministrativo. Quale spazio per la decisione algoritmica? - 8. *Privacy* e altre garanzie nel trattamento algoritmico dei dati. - 9. AI, algoritmi e “società dell’anticipazione”. Problemi del processo democratico. - 10. L’interrogativo “intrattabile”. Ci sarà una soggettività per gli AI systems o agents? - 11. Prime (e provvisorie...) conclusioni. Il diritto di fronte ad una sfida radicale e non temporanea.

1. Tra fiction e diritto. Alla ricerca di principi e regole per l’AI

Le questioni poste dall’AI costituiscono, almeno parzialmente, una delle frontiere nuove della bioetica e del biodiritto, portano quasi a ripensarne l’appropriatezza della dimensione settoriale rispetto ad un contenitore più largo, che potrebbe essere descritto sinteticamente come il luogo plurale del confronto tra sviluppi scientifici e tecnologici e scienze sociali (il diritto tra queste)¹.

* Una versione meno ampia di questo saggio è stata già pubblicata in *Biolaw Journal – Rivista di Biodiritto*, n. 1 del 2019.

** Università degli Studi di Parma.

1. Su «*AI Ethics as sub-field of applied ethics and technology*», cfr. *Draft Ethics Guidelines for Trustworthy AI*, predisposto dall’High Level Expert Group on Artificial Intelligence della Commissione Europea (d’ora in avanti DEG), del 18 dicembre 2018,

Si parla di “Roboetica” (o di AI ethics) come «ramo dell’etica applicata, cioè una riflessione filosofica, e perciò sistematica e informata, che studia le conseguenze tanto positive quanto negative dei robot (ma più in generale dei processi algoritmici e dei sistemi agenti autonomi) nella società allo scopo di sollecitarne la progettazione, lo sviluppo e l’uso morale»².

Al di là adesso di discussioni nominalistiche o di “regolamenti di confine”, il tema del rapporto tra AI e sistemi sociali e giuridici, fino a pochi anni fa limitato ad una sfera ristretta di specialisti, talvolta considerati anche un po’ visionari, oggi sembra esplodere in tutta la sua complessità di implicazioni e di sfide “adattive”.

Il tema dell’AI ha vissuto di fiammate e di delusioni (si è parlato di «primavere» e «inverni»)³. Dentro questi cicli molto oscillanti, è emerso ad ogni modo l’impatto su tutta una serie di aspetti della vita sociale ed economica, e hanno cominciato a prendere corpo interrogativi a volte capaci di scavare nel profondo dell’esperienza umana e di mettere in gioco basilari assunti etici ed antropologici.

E questo non poteva non chiamare in causa il diritto, la sua ricerca, spesso faticosa, di dare un senso alla realtà, allo scorrere delle vicende umane.

«Come regolare un mondo nuovo» non è allora solo un omaggio alla fantasia distopica di Aldous Huxley⁴. La sensazione è che effettivamente l’AI sta cambiando (o comunque appare capace di cambiare) radicalmente il mondo sotto i nostri occhi. Quello che fino a poco tempo fa sembrava dominio quasi esclusivo della *science fiction* (da Blade

2 e 6, che non a caso cita tra i documenti di riferimento di questo nuovo settore la Convenzione di Oviedo sulla Biomedicina del 1997, che è uno dei testi chiave della riflessione bioetica europea.

2. Così G. Tzafestas, *Roboethics. A navigating overview*, Springer, 2016.

3. Per una breve storia delle ricerche sull’AI, v. J. Bernstein, *Uomini e macchine intelligenti*, Milano, 1990, 19 ss., 54 ss., che ricorda la Conferenza estiva del Dartmouth College del 1956, e il Gruppo di Intelligenza Artificiale del MIT (Massachusetts Institute of Technology) creato nel 1958 da J. McCarthy e M. Minsky (quest’ultimo fu anche il consulente di Stanley Kubrick nella lavorazione di 2001-Odissea nello Spazio). A proposito delle delusioni, sul Teorema di Tesler, secondo cui «l’IA è tutto ciò che non è ancora stato fatto», v. D.R. Hofstadter, *Godel, Escher, Bach: un’eterna ghirlanda brillante*, trad. it., Milano, 1984, 649. Secondo C. Accoto, *Il mondo dato*, Milano, 2017, 9, invece, «la stagione dell’inverno dell’intelligenza artificiale potrebbe essere ormai definitivamente alle nostre spalle. Il lungo periodo di disaffezione fatto anche di fallimenti clamorosi, di carenza o scarsità di fondi e di interesse per la ricerca e lo sviluppo scientifico e industriale dell’intelligenza artificiale sembra si stia chiudendo».

4. Il riferimento è al celebre romanzo di A. Huxley, *Il Mondo nuovo* (1932). Di un nuovo mondo, a proposito dell’AI, parla L. Alexandre, *La guerra delle intelligenze*, trad. it., Torino, 2018, 26.

Runner a 2001 - Odissea nello spazio, fino ai più recenti Matrix, Transcendence, Her, Ex machina, The Circle, e le serie Westworld e Black Mirror⁵), sta assumendo forme concrete, quelle delle *driverless cars*, di sistemi diagnostici e/o chirurgici sempre più accurati e precisi (con livelli differenti di supervisione umana e «ad autonomia crescente»⁶), di algoritmi in grado di fare il *risk assessment* finalizzato a decidere se un soggetto deve essere mantenuto in stato di custodia cautelare, oppure può essere messo in libertà in attesa di accertare la sua responsabilità in un processo, o di fare lo *scoring* per la valutazione di ammissibilità di un contratto assicurativo o di mutuo, e altre cose ancora (nell'ambito ad esempio della sicurezza con il riconoscimento facciale, della medicina nell'analisi diagnostica delle immagini digitali, della finanza per la prevedibilità dell'andamento dei mercati, della gestione delle infrastrutture di mobilità o delle politiche ambientali, della determinazione dei contributi nell'ambito del Fondo Unico per lo Spettacolo o per l'assegnazione delle sedi agli insegnanti, da ultimo con riferimento alla valutazione del livello di gravità da attribuire alle Regioni per quanto riguarda le politiche di contenimento del contagio da Covid-19).

Invero, l'immaginario scientifico e letterario su queste possibilità tecnologiche è molto più risalente nel tempo rispetto agli esempi cinematografici prima richiamati. Basti pensare al Golem della leggenda praghese, o alla concezione cartesiana degli animali come "automi"⁷, e prima ancora ai racconti su automi nell'antico Egitto, in Cina, nella Grecia antica, all'Homunculus di Paracelso, alla testa parlante dei Templari, al "moderno Prometeo" di Mary Shelley, al leone "robotico" che Leonardo di Vinci regalò al re di Francia nel 1515, al turco che giocava a scacchi costruito da Wolfgang von Kempelen nella seconda metà del XVIII secolo⁸, al dibattito settecentesco (La Mettrie, D'Holbach, Helvétius)

5. Cfr., per una rassegna sintetica di questi titoli, S. Lenzi, G. Ciofalo, Wiener, *Goldrake e i Social Media. L'eredità della cibernetica tra scienza, tecnologia e comunicazione*, introduzione a N. Wiener, *La Cibernetica* (1948), Roma, 2017, 23-24.

6. Vedi infra, in questo volume, il saggio di D. Amoroso e G. Tamburrini, *La questione del controllo umano significativo sui sistemi robotici ad autonomia crescente*.

7. Criticata da Bernard de Fontenelle (avvocato ed aforista francese a cavallo tra '600 e '700) con la celebre riflessione sui cani e gli orologi (riportata in S. Hustvedt, *Le illusioni della certezza*, Torino, 2018, 191): «Dite che le bestie sono macchine così come lo sono gli orologi? Mettete un cane macchina e una cagna macchina fianco a fianco, e alla fine il risultato sarà una terza piccola macchina, mentre due orologi giaceranno fianco a fianco per tutta la vita senza mai produrre un terzo orologio».

8. Ne parla S. Hustvedt, nel suo splendido libro *Le illusioni della certezza*, Torino, 2018, 164 ss.

sul rapporto uomo-animale-macchina (accuratamente ricostruito da A. Punzi⁹), all'anatra di de Vaucanson¹⁰.

Cinema, letteratura, diritto. Più che una sequenza tra sistemi separati, sembra essere, sull'AI, una rincorsa a capire e a mettere a fuoco le proiezioni reali (secondo scale temporali differenti) del fenomeno. I mondi della "finzione" e dell'ordine giuridico si incontrano, collaborano, il primo mostra una straordinaria capacità di anticipare la realtà e gli scenari futuri¹¹ (aiuta a capire problemi che non possono essere studiati guardando il passato)¹², la "provocazione" diventa sempre più verosimile, fino al punto da entrare nell'edificio degli atti normativi¹³.

Parafrasando il Borges di *Tlon, Uqbar, Orbis Tertius*¹⁴, sembra quasi di vedere l'immagine – in quel racconto splendidamente descritta – di una realtà che comincia a cedere in più punti, di fronte ad un mondo illusorio e fantastico che improvvisamente comincia ad irrompere nel mondo (apparentemente?) reale. Anche Alessandro Baricco, nel suo *The Game*¹⁵, gioca con la contrapposizione/integrazione tra mondo

9. Nel suo *L'ordine giuridico delle macchine*, Torino, 2003.

10. Ne parla Gerard M. Edelman, *Sulla materia della mente*, Milano, 1993, 293.

11. Come scrivono V. Gallese, M. Guerra, *Lo schermo empatico. Cinema e neuroscienze*, Milano, 2015, 85, «il cinema è visione del mondo, in tutta la sua composita ampiezza e in tutta la sua affascinante imprevedibilità, visione del mondo che allarga i nostri orizzonti percettivi, cognitivi, immaginativi e che si cala talmente a fondo nel dialogo con il suo tempo e con la storia da restituircene un'interpretazione e una modellazione incredibilmente profonda e completa, che va dal nostro rapporto con la tecnica alla nostra vita sociale, culturale e politica».

12. Cfr. A. Maynard, *Films from the future. The technology and morality of sci-fi movies*, Miami, 2018.

13. Secondo J. Balkin, *The Path of Robotics Law*, 60, «these literary and artistic sources ... already give us a sense of what robots might do and the problems that their introduction into society might cause». Più in generale, per S. Almong, *Literature alongside law as contemporary paradigm*, in *Cultural Dynamics*, 2001, 59, «both law and literature (are) social practices in constant interaction. It contends that aesthetics and poetics are an inextricable part of law, just as legal and ethical structures are an essential part of aesthetic expression. Narrative, as a basic component of human experience, continually provide one of the formative as well as change-inducing forces of cultural activity, including legal activity».

14. Pubblicato in *Finzioni*, ora in *Tutte le opere*, vol. I, Milano, 1984, 623-641.

15. A. Baricco, *The Game*, Torino, 2018, 88-89, e 191, dove si legge: «il traffico con l'oltremondo, in entrata e in uscita, è diventato immenso e velocissimo, tanto immenso e tanto veloce che spesso conservare una vera linea di demarcazione tra mondo e oltremondo è diventato impossibile, e quasi sempre inutile». Cfr. anche, sulla giustapposizione tra mondo reale e mondo digitale ("datasfera", «fatta di miliardi di miliardi di dati») V. Zeno-Zencovich, *La "datasfera". Regole giuridiche per il mondo digitale parallelo*, in L. Scaffardi (a cura di), *I "profili" del diritto. Regole, rischi e opportunità nell'era digitale*, Torino, 2018, 99 ss.

“reale” e oltremondo digitale (la “infosfera”¹⁶ di cui parla L. Floridi). Per Baricco, l’habitat dell’uomo digitale «è un sistema di realtà a doppia forza motrice, dove la distinzione tra mondo vero e mondo virtuale decade a confine secondario, dato che l’uno e l’altro si fondono in un unico movimento che genera, nel suo complesso, la realtà. [...] è un sistema in cui mondo e oltremondo girano uno nell’altro producendo esperienza, in una sorta di creazione infinita e permanente, [...] ricamando una trama che legittimamente chiamiamo REALTÀ»¹⁷.

Il ‘gioco’ letteratura-diritto funziona altresì in senso inverso, nel senso cioè che il diritto si appropria di immagini della letteratura, prova a svolgere la sua funzione partendo da queste suggestioni.

Emblematico il richiamo che la Risoluzione del Parlamento Europeo del 16 febbraio 2017 contenente *Norme di diritto civile sulla Robotica* fa, in uno dei suoi *Considerando*, alle leggi di Asimov¹⁸ come base di riferimento etica e deontologica dei comportamenti dei ricercatori che operano nel settore della robotica, e del loro impegno a rispettare i principi di beneficenza (secondo cui i robot devono agire nell’interesse degli

16. L. Floridi, *Pensare l’infosfera. La filosofia come design concettuale*, Milano, 2020. C. Accoto, *Il mondo ex machina*, Milano, 2019, 4, parla di «un mondo automato che urge e che merita di essere analizzato in sé, senza preconcetti neoluddisti, ma con apertura, consapevolezza e profondità».

17. C. Accoto, *Il mondo dato*, Milano, 2017, 5, parla di un «Mondo dato», espressione che «ha, al tempo stesso, una valenza poetica e un intento provocatorio. È poetico nella misura in cui gioca, linguisticamente, con un aggettivo (il mondo come realtà data, come entità chiusa) evocando però anche il sostantivo (il mondo-dato, il mondo come entità quantificata, fatto di e dai dati)». Su una “natura digitale delle cose”, come ordine spontaneo in cui «un insieme di gruppi sociali [...] si connettono per via evolutiva, e competitiva, dandosi regole private che cessano però di essere squisitamente contrattuali, divenendo anche sociali e istituzionali», v. A. Venanzoni, *Intersezioni costituzionali – Internet e Intelligenze artificiali tra ordine spontaneo, natura delle cose digitale e garanzia dei diritti fondamentali*, in *Forum di Quaderni Cost.*, 27 aprile 2018, 8.

18. Nel Considerando T si legge: «considerando che le leggi di Asimov devono essere considerate come rivolte ai progettisti, ai fabbricanti e agli utilizzatori di robot, compresi i robot con capacità di autonomia e di autoapprendimento integrate, dal momento che tali leggi non possono essere convertite in codice macchina». Le famose leggi di Asimov sono le seguenti: «(1) Un robot non può recar danno a un essere umano né può permettere che, a causa del proprio mancato intervento, un essere umano riceva danno. (2) Un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non contravvengano alla Prima Legge. (3) Un robot deve proteggere la propria esistenza, purché questa autodifesa non contrasti con la Prima o con la Seconda Legge (cfr. Isaac Asimov, *Circolo vizioso*, 1942), e (0) Un robot non può recare danno all’umanità, né può permettere che, a causa del proprio mancato intervento, l’umanità riceva danno».

esseri umani), di non malvagità (in virtù della quale i robot non devono fare del male ad un essere umano), di autonomia (vale a dire la capacità di adottare una decisione informata e non imposta sulle condizioni di interazione con i robot), di giustizia (nel senso di un'equa ripartizione dei benefici associati alla robotica, e di accessibilità economica ai robot addetti all'assistenza a domicilio e alle cure sanitarie¹⁹).

C'è stato poi molto altro dopo questa Risoluzione. Maria Zanichelli ricostruisce con attenzione quella che chiama la «via europea all'IA»²⁰, che – come nota la stessa Autrice – «aspira a combinare l'implementazione dei sistemi intelligenti, quale fattore di innovazione, di profitto e di progresso a vari livelli, con l'attenzione ai rischi e alle criticità che essi implicano [...] (dentro) un quadro etico-giuridico chiaro e riconoscibile, capace di affrontare l'IA *in primis* in termini di sostenibilità, di responsabilità, di diritti delle persone, di vantaggi per la società, di attendibilità e trasparenza dei processi decisionali».

Si tratta essenzialmente di *soft law* che si combina con strumenti di self regulation (mettendo da parte ovviamente le disposizioni del GDPR)²¹. È inevitabile quando ci si pone di fronte a tecnologie “emergenti”. Il diritto procede in modo “progressivo”, sussidiario; ma questo, come è stato sottolineato giustamente, «può rappresentare l'approccio migliore per affrontare problemi complessi e diversi caratterizzati da incertezza»²².

19. Come sottolinea infatti A. Baricco, *The Game*, cit., 320, «... il *Game* è difficile. Magari divertente, ma troppo difficile. [...] Per sopravviverci bisogna avere skill non indifferenti che peraltro non vengono insegnate. [...] Non ci sono reti di protezione, né sistemi per recuperare chi è caduto. [...] Non lasceremo nessuno indietro non è una frase da *Game*».

20. M. Zanichelli, *Ecosistemi, opacità, autonomia: le sfide dell'Intelligenza artificiale in alcune proposte recenti della Commissione europea*, infra, p. 70. Da ultimo v. il documento *Getting the future right. Artificial Intelligence and fundamental rights*, pubblicato dalla European Union Agency for Fundamental Rights il 14 dicembre scorso.

21. Sollecita un approccio diverso G. Resta, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di uguaglianza*, in *Pol. Dir.*, n. 2/2019, 199 ss., secondo cui «la strada più proficua appare quella dell'intervento a geometria variabile, composta cioè da forme più soft di incentivazione all'adozione di tecnologie e prassi organizzative human rights compliant [...] e strumenti più incisivi con funzione prettamente regolamentare [...] i quali dovrebbero poi ricadere a cascata sulla fase della programmazione dell'algoritmo».

22. E. Stradella, *La regolazione della Robotica e dell'Intelligenza artificiale: il dibattito, le proposte, le prospettive. Alcuni spunti di riflessione*, in *Media laws*, n. 5/2019, 78-79. Vedi anche G. Guerra, *La sicurezza degli artefatti robotici in prospettiva comparatistica*, Bologna, 2018, 57-58.

Da questi Documenti cominciano ad emergere tutto un fascio di problemi e di possibili connessioni/conflitti degli sviluppi dell'AI con diritti e principi giuridici fondamentali.

Le “leggi” di Asimov prima richiamate appaiono semplicemente una primissima (e parziale) impostazione del discorso sui principi etici e giuridici dell'AI. Il quadro da costruire è assai più complesso, e di questa complessità si avverte una consapevolezza via via più profonda nei documenti “paranormativi” (*guidelines, statement, studi preparatori*) elaborati ai più vari livelli istituzionali. Si lavora essenzialmente in chiave di adattamento interpretativo dei principi consolidati del costituzionalismo alle applicazioni di AI: l'obiettivo è quello di configurare un'AI “*human-centric*” and “*trustworthy*”, appunto perché connotata da un “*ethical purpose*”, e da un obbligo di conformità ai valori fondamentali della convivenza civile (rispetto della dignità umana e dei diritti, eguaglianza e non discriminazione – come “*fairness*”, nel duplice significato di difesa e prevenzione “*from unfair bias*”, e di eguale opportunità di accesso a queste possibilità tecnologiche, alle risorse del mondo digitale, ricostruibili come “beni comuni”²³ o come oggetto di un vero e proprio diritto sociale²⁴ –, non interferenza rispetto ai processi democratici, sicurezza, rispetto per la *privacy*), e al benessere delle persone²⁵.

La dignità umana viene tradotta, in un significato relazionale, come diritto delle persone umane di sapere *se* e *quando* stanno interagendo con una macchina o con un altro essere umano, e di decidere *se, come* e *quando* attribuire determinati compiti ad un sistema artificiale autonomo o ad una persona. L'autonomia delle persone a contatto con AI

23. Su Internet e sui nuovi beni comuni digitali, v. G. Mulgan, *Big Mind. L'intelligenza collettiva che può cambiare il mondo*, trad. it., Torino, 2018, 239 ss., 252.

24. Vedi T.E. Frosini, *Il costituzionalismo della società tecnologica*, in *Dir. dell'informazione e dell'informatica*, n. 4/2020, 470.

25. Sul punto, v. DEG, cit., iv. A questi principi di riferimento è legato anche il criterio secondo cui «*AI systems should be developed and implemented in a way that protects societies from ideological polarization and algorithmic determinism*» (cit., 9); in sintesi, secondo il documento citato, una AI “*trustworthy*” richiede una serie di requisiti, tra cui: 1. *Accountability*, secondo diversi meccanismi di compensazione o di “*reconciliation*” anche non monetari; 2. *Data governance*, in modo da evitare o prevenire il più possibile biases, errori, effetti discriminatori, o da migliorare i sistemi attraverso la ricerca e la verifica degli errori; 3. *Design for all*, nel senso di massima accessibilità ed usabilità di questi *tools*, senza riguardo a condizioni personali, età, disabilità; 4. *Governance of AI autonomy (Human Oversight)*, commisurata in modo crescente proprio al grado di autonomia operativa delle macchine; 5. *Respect for privacy*, in totale compliance con le norme del GDPR; 6. *Robustness*, intesa come sicurezza degli algoritmi durante le diverse fasi di implementazione, e resilienza rispetto ad eventuali attacchi informatici.

agents o systems deve comportare altresì che i meccanismi di funzionamento di questi sistemi siano trasparenti e prevedibili, meglio ancora “esplicabili”, «comprehensible and intelligible by human beings at varying levels of comprehension and expertise»²⁶ (su questo torneremo anche più avanti).

Il bene primario dell'integrità e della sicurezza personale assume una connotazione anche “emozionale”, mentale, legata all'interazione uomo-macchina, e agli effetti che le diverse modalità che questa interazione può assumere possono riflettersi sull'atteggiamento delle persone, sul modo di considerare i sistemi agenti in particolare quando adottano “fisicità” antropomorfe, sulle influenze distorsive che si possono produrre anche sul piano emotivo, soprattutto per le persone che si trovano in una condizione di debolezza e di vulnerabilità.

Infine, rispetto agli sviluppi dell'AI viene richiamata (sempre nello Statement del Gruppo di esperti della Commissione Europea) un'esigenza di sostenibilità (Simoncini preferisce il concetto più *hard* di precauzione costituzionale), di mantenimento nel tempo (e di garanzia per generazioni future) delle condizioni di base della vita umana e della natura: un'ulteriore conferma, se ce ne fosse bisogno, della preoccupazione che accompagna l'evoluzione di questa tecnologia, pur nella convinzione delle tante opportunità positive che da essa possono derivare.

2. AI, Robotics, mondo “digitale”

Ad ogni modo, l'AI sembra davvero qualcosa in più di un set di risorse tecnologiche aggiuntive e avanzate, con le quali fare semplicemente meglio e in modo più rapido ed efficace quello che possiamo già fare²⁷.

Non è semplicemente un “mezzo”, ma un “mondo”, per dirla con Gunther Anders: «e mondo è qualcosa di diverso da “mezzo”, appartiene

26. Cfr. DEG, cit., 10 ss.; in particolare, secondo questo documento, «*xplicability also requires accountability measures be put in place. Individuals and groups may request evidence of the baseline parameters and instructions given as inputs for AI decision making [...] by the organizations and developers of an AI system, the technology implementers, or another party in the supply chain*».

27. Come scrive B. Reese, *La quarta era. Robot intelligenti, computer consapevoli e il futuro dell'umanità*, Milano, 2019, 39, «il computer non è semplicemente un gadget, ma un dispositivo che ha una rilevanza filosofica». Cfr. anche L. Floridi, *La rivoluzione dell'informazione*, Torino, 2012, 13, secondo cui «*le ICT non stanno soltanto ricostruendo il mostro mondo: lo stanno riontologizzando*».

a una categoria diversa»²⁸. Si parla di “tecnologie profonde”, “invenzioni che ci reinventano”, che «segnano un genere diverso di cambiamento del mondo» proprio perché puntano a «progettare meglio l’evoluzione»²⁹.

Alcuni aspetti di questo imponente fenomeno ci pongono davanti ad interrogativi etici ed antropologici che rappresentano un vero e proprio cambio di paradigma³⁰, al fondo del quale potrebbe esserci una nuova idea (e una nuova esperienza) dell’umanità³¹.

Inoltre, come è stato acutamente sottolineato, l’AI non è un problema temporaneo, non è qualcosa che ad un certo punto sparisce, ma è

28. G. Anders, *L’uomo è antiquato. I: Considerazioni sull’anima nell’epoca della seconda rivoluzione industriale*, Torino, 2003, 38. Molto acutamente L. Floridi, *Why information matters*, in *The New Atlantis*, special issue, winter 2017, sottolinea: «*Information technologies don’t just modify how we act in the world; they also profoundly affect how we understand the world [...] information needs to be a significant field of philosophical study [...] We need a philosophy of information*». Sempre L. Floridi, *Pensare l’infosfera*, cit., 130, sottolinea che «*nel presente, stiamo lentamente accettando l’idea che non siamo enti isolati ed unici, quanto piuttosto organismi, il cui sostrato è informazionale (infosfera), reciprocamente connessi e parte di un ambiente costituito da informazioni (infosfera), che condividiamo con agenti naturali o artificiali simili a noi sotto più profili. Turing ha cambiato la nostra antropologia filosofica tanto quanto Cartesio, Darwin o Freud*». In senso analogo, v. le riflessioni di C. Accoto, *Il mondo ex machina*, cit., 6-7, che parla dell’automazione «non come meccanismo neo-macchinico, ma come dinamismo neo-ecologico», sottolineando l’esigenza di un approccio “philtech” che prenda atto del fatto che «il mondo si è avviato ... a una trasformazione profonda e irreversibile delle proprie fondamenta ontologiche e ontogenetiche».

29. Per queste definizioni teoriche v. ora J. Preston, *L’era sintetica*, Torino, 2019, 175 ss., 52, 62. Cfr. anche U. Galimberti, *Psiche*, 345 ss., che parla di una tecnica «*che decide il modo di percepire, di sentire, di pensare e di progettare dell’uomo*», «*che dispone della natura come suo fondo e dell’uomo come suo funzionario*»; e ancora (521), secondo Galimberti, ci troviamo in un passaggio epocale, in cui «*si tratta di scoprire cosa l’uomo diventa nell’orizzonte della sperimentabilità illimitata e della manipolabilità infinita disciuso dalla tecnica*».

30. A. Punzi, *L’ordine giuridico delle macchine*, Torino, 2003, 18.

31. Nelle *Draft Ethics Guidelines for Trustworthy AI*, predisposto dall’*High Level Expert Group on Artificial Intelligence* della Commissione Europea (d’ora in avanti DEG), del 18 dicembre 2018, si legge che «*Artificial Intelligence is one of the most transformative forces of our time, and is bound to alter the fabric of society*». In termini, v. anche F. De Vanna, *Diritto e nuove tecnologie: il nodo (controverso) della regolazione giuridica*, in *Lo Stato*, 2018, n. 11, 388; e V. Zeno Zencovich, *Big data e epistemologia giuridica*, in S. Faro, T.E. Frosini, G. Peruginelli (a cura di), *Dati e algoritmi. Diritto e diritti nella società digitale*, Bologna, 2020, 13, secondo cui «*la questione è se questo scenario tecnologico relativamente recente, che sta crescendo ad un ritmo impressionante e che, apparentemente, non può essere fermato, cambia i tradizionali modelli epistemologici del giurista, ovvero il modo in cui comprende il mondo ed offre strumenti giuridici per il suo miglior funzionamento*».

una novità destinata a provocare una fase nuova dell'esperienza umana con la quale dovremo abituarci a convivere nel tempo³².

Intanto, non è facile stabilire cosa è AI³³; manca una definizione univoca e generalmente accettata, e soprattutto cambia nel tempo³⁴, il che è normale per una tecnologia “generativa”, che «changes over time as people work with and through new technology»³⁵.

Di certo, non è solo robotica³⁶. Quando parliamo di sistemi di AI, può esserci un “corpo” (vale a dire un contenitore fisico, anche variamente antropomorfo) oppure no: questi sistemi possono lavorare in uno spazio fisico ovvero in uno spazio virtuale.

È possibile pensare a pezzi di AI anche dentro di noi (interfacce neurali, *Brain computer interfaces*), come risorse di potenziamento cognitivo o fisico-biologico; e i dilemmi non sono meno urgenti e difficili di quelli che riguardano l'*enhancement* genetico e neurobiologico³⁷. Si ipotizzano altresì meccanismi di *mind reading*³⁸ o *mind uploading* capaci di estrarre informazioni dal cervello umano e di replicarle in un computer³⁹.

32. Cfr. L. Alexandre, *La guerra delle intelligenze*, cit., 295, sottolineando che «l'IA ormai non è più un'opzione disattivabile, un interruttore che avremo ancora la possibilità di spegnere. È diventata indispensabile». Sempre questo A. ha messo in evidenza che il cervello (tra neuroscienze e AI) è diventato il principale campo di battaglia di questo secolo.

33. D'altronde, come spiega E. Di Mauro, *La mente naturale e la mente artificiale*, Trieste, 2019, 19, anche sulla definizione di intelligenza tout court esistono centinaia di risposte differenti.

34. R. Angelini, *Intelligenza artificiale e governance. Alcune riflessioni di sistema*, in *Astrid*, n. 14/2017, 1. Una definizione ampia si legge in DEG, iv: «Artificial Intelligence refers to systems designed by humans that, given a complex goal, act in the physical and digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to learn to adapt their behaviour by analysing how the environment is affected by their previous actions».

35. Per questa definizione di “generative technology”, cfr. J. Balkin, *The Path of Robotics Law*, 47.

36. M. Luciani, *La decisione giudiziaria robotica*, in *Nuovo diritto civile*, n. 1/2018, 7.

37. Sul tema, sia consentito il rinvio ad A. D'Aloia, *Oltre la malattia: metamorfosi del diritto alla salute*, in *Biolaw Journal*, n. 1/2014, 87 ss., 93 ss.

38. Per alcune prime e parziali (rispetto alla portata potenziale della prospettiva) applicazioni di tecniche di *mind reading* all'interno di un processo, basate essenzialmente sull'uso dell'Implicit Association Test per verificare/confermare l'immagine che la persona aveva conservato di un determinato evento, v. S. Amato, *Biodiritto 4.0*, Torino, 2020, 144.

39. S. Rodotà, *Il diritto di avere diritti*, cit., 347. U. Ruffolo, A. Amidei, *Intelligenza artificiale*, 208-209, analizzano alcune davvero inquietanti prospettive di queste tecni-

Ma cosa (o chi) sarà alla fine di questo processo? Se si dovesse arrivare davvero alla possibilità di trasposizione dei contenuti mentali in un “luogo” (contenitore? Identità? Individuo?⁴⁰) digitale, quale sarà l’esito di tutto questo? Cosa resterà della condizione umana o naturale di fronte ad un umano potenziato e collegato a molteplici forme di intelligenza collettiva⁴¹?

Insomma, i “due mondi”, “naturale” e “artificiale”, non si limitano a contrapporsi ma si incrociano, si sovrappongono, sperimentano prospettive di ibridazione in cui i confini si scolorano⁴², perdono nettezza, e la pretesa e stabile “datità” della natura viene messa in discussione dalla capacità conformativa e manipolatoria della tecnica⁴³.

che, come il mind uploading, o il trapianto di cervello (attualmente vietato dall’art. 3 della l. 91/1999), affermando, con intelligente provocazione: «È fantadiritto? Forse solo (ancora) surreale. Perché siamo già fuori dalla fantascienza, che comunque, e non di rado, è lo stato provvisorio di previsioni che si realizzano in scienza [...]. Si tratta di traguardi non ancora raggiunti, né forse prossimi, ma che la scienza si prefigura, e che già concretamente studia in alcuni centri d’avanguardia, perseguendo sogni e incubi già preconizzati come possibile realtà futura, e dunque come terreno di mediazione giuridica fino a ieri inconcepibile, ma in relazione al quale riflettere proattivamente per il domani». Per una citazione cinematografica dell’ipotesi della replicazione del cervello, e dello “sparpagliamento” di diversi avatar, v. S. Arcagni, *L’occhio della macchina*, Torino, 2018, 169-170.

40. Vedi anche S. Amato, *Biodiritto 4.0*, cit., 148 e 153, dove questo studioso sottolinea acutamente un collegamento di questo tema con il divieto legislativo di trapianto del cervello, quasi a segnare «una soglia ideale di inviolabilità della specificità personale [...] da cui ricavare il divieto di ogni copyright sul cervello in quanto tale e su ogni singola manifestazione del cervello».

41. Come ipotizza G. Mulgan, *Big Mind*, cit., 261 ss., 270, affermando (a pag. 63) che, quando questo sarà possibile, «l’idea che il pensiero abbia luogo esclusivamente nella mente di individui autonomi e isolati apparirà sempre più peregrina». Di questa prospettiva di un “brain net” parla anche S. Amato, *Biodiritto 4.0*, cit., 146.

42. Scrive A. Punzi, *L’ordine giuridico delle macchine*, Torino, 2003, «i codici della vita sembrano destinati ad incontrarsi con i codici della materia».

43. Per S. Maffettone, *Il valore della vita*, Milano, 1998, 60, «la natura non dice che cosa è naturale e che cosa non lo è». Splendida questa riflessione ancora di N. Irti, *L’uso giuridico*, cit., 19: «Anche la natura non è più la stessa. [...] Il pensiero moderno [...] incalza contro l’immota oggettività della natura. La quale non ci sta più dinanzi con la perentoria richiesta di adeguarci ad essa, con l’alterigia delle cose che non nascono e non muoiono, ma è un oggetto di pensiero, fino a diventare un oggetto posto dal pensiero». Sempre Irti aggiunge: «l’artificialità si allarga anche alla natura, vista non più come il “naturale”, dato da sempre e per sempre, contrapposto all’effimero e precario “artificiale”, ma conformata, anch’essa, dall’uomo, che, indagandola con propri occhi e strumenti d’osservazione, incontra dovunque se stesso. Neppure la natura può sottrarsi all’umanità, ossia al volere e al pensiero che le imprimono forma e la fanno la “nostra natura”. [...] Allora, se umanesimo e tecnica si vogliono in antitesi, non sarà più opposizione tra umano e non umano, ma all’interno stesso dell’umano».

Più concretamente, almeno per ora, le BCI possono avere applicazioni più immediate e prive dei risvolti inquietanti immaginati a proposito del mind reading⁴⁴, come ad esempio la possibilità di mantenere uno spazio di autonomia motoria per pazienti tetraplegici⁴⁵, ovvero di autodeterminazione e di espressione alternativa della volontà per soggetti gravemente malati sul piano mentale, e non in grado di comunicare nemmeno con il movimento di parti del corpo (pensiamo ai casi di locked-in syndrome, o ai casi di SLA avanzata). E questi sono certamente risultati apprezzabili, prospettive da incoraggiare.

Le sperimentazioni di Adrian Owen sulla c.d. “grey zone” (relativa alla condizione dei pazienti in stato vegetativo o in stato di minima coscienza) mostrano che è possibile lavorare su risposte secche a domande precise, istruendo l’interessato a pensare in un certo modo, come ad esempio camminare nella propria casa, o alzare una mano, giocare a tennis o sollevare un piede. Un meccanismo analogo può essere utilizzato per selezionare un vocabolo, e così, faticosamente, comporre parole, frasi, ovviamente semplici.

In generale, però, le potenzialità di penetrazione della tecnica nella sfera più intima del soggetto, quella del pensiero e dei suoi correlati neurali, solleva interrogativi di natura etica e giuridica su come proteggere questi dati così “sensibili”, anche dai rischi di hackeraggio e di una circolazione potenzialmente discriminante; sui confini e sulle implicazioni della libertà (e della continuità della storia) psichica e cognitiva di un soggetto, della sua “integrità” mentale; sulla possibilità – che deve essere sempre garantita – di disconnettersi dal mezzo di interfacciamento.

In un recentissimo documento del Consiglio d’Europa (Risoluzione n. 2344 del 22 ottobre 2020), queste preoccupazioni sono espresse senza molti giri di parole, a cominciare dal titolo, dove c’è un interrogativo sulle potenzialità finalistiche delle brain computer interfaces: «new rights or new threats to fundamental freedoms»?

Nel documento ci si sofferma non solo sui rischi di violazione della privacy e della dignità, «with the potential to subvert free will and breach the ultimate refuge of human freedom»; ma altresì sul pericolo distorsivo che si annida nell’uso “potenziativo” delle BCI, che potrebbe creare «separate categories of human beings, the enhanced and the unenhanced», interferendo in senso diminutivo sull’identità individuale, sull’agire umano e sulla responsabilità morale⁴⁶.

44. Per le quali J. Kaplan, *Intelligenza artificiale*, Roma, 2016, 215, conclude dicendo «spero che queste preoccupazioni non diventino mai le mie».

45. Vedi M. Cobb, *The Idea of the Brain. A History*, London, 2020, 288 ss.

46. Nel documento si arriva a dire che «Such outcomes could change the very na-

Ad ogni modo, l'incertezza definitoria che abbiamo visto in generale per l'AI, vale anche per la parola e la nozione di "robot" in senso stretto. Come scrive J. Engelberger, «*I can't define a robot, but I know one when I see one*»⁴⁷. In altri termini, anche un robot, che a sua volta è una proiezione parziale del più ampio *field* dell'AI, può essere tante cose e può assumere diverse configurazioni tassonomiche, che dipendono da cosa è chiamato a fare, dall'"ambiente" di lavoro, dalle tipologie di interazione con gli esseri umani, dai gradi o livelli di autonomia decisionale e/o operativa⁴⁸.

Dunque, AI appare effettivamente un "*umbrella term*", una parola "contenitore", che include – si può dire in via di prima approssimazione – una varietà di tecniche computazionali e di processi associati (di tipo algoritmico) dedicati a migliorare l'abilità delle macchine nel fare cose che richiedono intelligenza⁴⁹.

Sullo sfondo, e anzi al cuore di questo processo, scorre un flusso inesauribile e gigantesco di dati⁵⁰.

ture of humanity and of human societies". L'Assemblea del Consiglio d'Europa indica una serie di principi etici da applicare allo sviluppo e all'uso delle neurotecnologie in generale e delle BCI in particolare. Tali principi possono essere così riassunti: 1) *Beneficence and prevention of malign use*; 2) *Safety and precaution*; 3) *Privacy and confidentiality*; 4) *Capacity and autonomy*; 5) *Human agency and responsibility*; 6) *Equity, integrity, and inclusiveness*; 7) *Ensuring public trust through transparency, consultation and education/awareness-raising*. Viene infine prospettata l'opportunità di una riflessione specifica sui nuovi "*neurorights*", ipotizzando anche l'adozione di un Protocollo aggiuntivo alla Convenzione EDU.

47. J. Kaplan, *Intelligenza artificiale*, cit., 30.

48. P. Dumouchel, L. Damiano, *Vivere con i robot. Saggio sull'empatia artificiale*, Milano, 2019, 167 ss.

49. Per questa definizione, v. il Report di Aa.Vv., *Artificial Intelligence & Human Rights: Opportunities and Risks* (d'ora in avanti AIHR), Berkman Klein Center for Internet & Society at Harvard University, 25 sept 2018, 10, dove si sottolinea altresì che «*what is considered artificial intelligence changes over time*». Cfr. anche European Commission for the efficiency of Justice (CEPEJ), *European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*, 3-4 dicembre 2018, 31, dove si legge che «*AI is not a single, homogeneous object: it is actually an assemblage of sciences and techniques (mathematics, statistics and computer science) capable of processing data to design very complex computer processing tasks*».

50. Vedi M. Du Sautoy, *Il codice della creatività*, Milano, 2019, 88, il quale fornisce un dato in sé emblematico, e cioè che «oggi l'umanità produce in due anni la stessa quantità di dati che abbiamo generato dall'alba della civiltà al 2003». Aggiunge questo autore che «*questa marea di dati è il catalizzatore principale della nuova era dell'apprendimento delle macchine. [...] chi ha accesso ai dati è seduto a cavalcioni sui pozzi di petrolio del XXI secolo*». La metafora dati/petrolio è usata anche da F. Foer, *I nuovi poteri forti*, Milano, 2017, 215, secondo cui i dati (quello che compriamo, la nostra