

Persona elettronica. Una finzione giuridica per l'intelligenza artificiale

Paolo Moro

Università degli Studi di Padova

Abstract: Electronic Person. A Legal Fiction for Artificial Intelligence

In the premise, it is observed that the technological advancement of artificial intelligence is pushing advanced mechatronic systems to transform into sophisticated robotic forms, which now lack an android morphology but emulate typically human faculties and resemble artificial persons. The introductory part of the study critically explores the foundational aspect of robotic subjectivity and re-examines the epistemological parallels, commonly accepted in the modern era, between the calculating machine and human consciousness: the traditional comparison between the digital device and human reasoning capability; the view of software as agents, carriers of ethical and legal responsibility; the definition of mechanical entities capable of recognising and emulating emotions. In the next part of the essay, the tendency to superficially humanise machines, through a naive imitation of human behaviour, is criticised, and the theoretical and practical difficulties of establishing an acceptable comparison between artificial intelligence and human consciousness are highlighted, still presupposed on the basis of a reductive conception of human subjectivity. In the conclusion, the question is raised as to whether and to what extent, based on the epistemological premises outlined in the first part, sophisticated mechatronic systems and advanced robotic entities might attain the qualification of “electronic person”, a legal fiction indicated by the Resolution of the European Parliament of 16 February 2017 and currently overlooked by the recent AI Act of the European Union.

Keywords: Robotic Subjectivity, Electronic Person, Legal Fiction, Mechatronic Systems and Law, Moral Machines.

Sommario: 1. Persone artificiali – 2. Congegni intelligenti – 3. Congegni morali – 4. Congegni emotivi – 5. Congegni inconsapevoli – 6. Persone elettroniche.

1. Persone artificiali

La realizzazione di sistemi di intelligenza artificiale altamente organizzati, dotati di elevata automazione e capaci di emulare alcune facoltà tipicamente umane, non è più il frutto della fervida immaginazione letteraria della fantascienza più brillante e

riflessiva del Novecento¹, ma un fenomeno sociale ed economico che ha provocato in Europa molteplici studi sui limiti etici dei sistemi meccatronici e cibernetici e ha condotto all'approvazione del nuovo AI Act dell'Unione Europea sull'intelligenza artificiale, che nell'intero pianeta si presenta come la prima legge che disciplina il fenomeno.

Come è noto, l'ampia e progressiva diffusione dell'intelligenza artificiale, alimentata dai rapidi progressi della tecnologia informatica, ha portato l'Unione Europea a considerare l'importanza del tecnodiritto² e stabilire una nuova legislazione cogente al fine di garantire lo sviluppo continuo del settore industriale e di proteggere simultaneamente il valore intrinseco della persona in base a principi generali di tutela della dignità umana³.

Pertanto, dopo una serie articolata di lavori preparatori, il Parlamento Europeo ha approvato definitivamente il 13 marzo 2024 il regolamento sull'intelligenza artificiale (*artificial intelligence act*), che è un atto normativo di portata generale e, quindi, è obbligatorio in tutti i suoi elementi e direttamente applicabile in ciascuno degli Stati membri (in base all'art. 298 del Trattato sul Funzionamento dell'Unione Europea).

Già una decina d'anni orsono si avvertiva la possibilità che, attraverso l'utilizzo più efficace dell'intelligenza artificiale, i nuovi saperi del terzo millennio, come l'informatica, la robotica e le neuroscienze, avrebbero condotto alla costruzione di sistemi meccatronici complessi, che appaiono quasi come persone artificiali, in quanto imitano differenti facoltà umane: non soltanto ragionamenti, come la comprensione del linguaggio naturale, l'apprendimento automatico o la rappresentazione della conoscenza, ma anche comportamenti, come la capacità di decidere o la reazione a stimoli esterni, e in taluni casi anche emozioni, attraverso le espressioni del volto sintetico o altre azioni comunicative⁴.

È arrivata con inattesa sollecitudine l'era delle macchine come noi⁵: come è già stato illustrato in un precedente lavoro⁶, l'informatica e la robotica permettono oggi di costruire *robot* anche immateriali che possiedono non solo l'abilità di calcolare, ossia di elaborare dati in base al modello computazionale come ogni *computer*, ma anche la capacità di decidere, attraverso la retroazione (*feedback*) o

¹ P.K. Dick, *Do Androids Dream of Electric Sheep?*, Doubleday and Company, New York, 1968.

² P. Moro, C. Sarra (a cura di), *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, FrancoAngeli, Milano, 2017.

³ L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, E. Vayena, "AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations", in *Minds & Machines*, 28 (2018), n. 4, pp. 689-707.

⁴ P. Moro, "Biorobotics and fundamental rights. Issues and limits of artificial intelligence", in D. Provolo, S. Riondato, F. Yenisey (eds.), *Genetics Robotics Law Punishment*, Padova University Press, Padova, 2014, pp. 517-531.

⁵ I. McEwan, *Machines like me*, Jonathan Cape, London, 2019.

⁶ P. Moro, "Macchine come noi. Natura e limiti della soggettività robotica", in U. Ruffolo (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Giuffrè, Milano, 2020, pp. 45-62.

l'apprendimento automatico (*machine learning*), e anche, in certe condizioni, la facoltà di manifestare percezioni simili alle emozioni.

Tuttavia, i problemi etici e giuridici posti dall'intelligenza artificiale, talora erroneamente vista come un apparato univoco o un dispositivo indipendente, riguardano principalmente l'interazione umana con i sistemi robotici evoluti, i quali appaiono sempre più capaci di emulazione autonoma delle facoltà proprie del soggetto umano, del quale influenzano il modo di vivere ma dal cui controllo restano sempre più emancipati⁷.

Questa rappresentazione meccanica dell'anima umana, riflessa come in uno specchio da un sistema non soltanto materiale (*hardware*) ma anche immateriale (*software*), sempre più simile alla persona e sempre più capace di interagire con essa, ripropone la questione filosofica del discusso paragone tra coscienza umana e intelligenza artificiale e ripropone temi e problemi di etica, legalità e tecnologia, che sono studiati e discussi nell'informatica giuridica contemporanea⁸.

In particolare, non è peregrino ribadire che la macchina intelligente, sempre che sia dotata di facoltà artificiali analoghe, se non superiori, a quelle umane, possa essere titolare di una sorta di autodeterminazione individuale e, quindi, anche di posizioni giuridiche soggettive simili a diritti e doveri: infatti, la capacità di apprendimento del sistema evoluto di intelligenza artificiale è interattiva e aumenta in modo progressivo quando opera con l'essere umano⁹.

Un esempio plastico di questa interazione è l'architettura tecnologica dei più evoluti *chatbot*, ossia di programmi cibernetici che realizzano assistenti virtuali progettati per simulare conversazioni umane e utilizzati non solo per supporto tecnico o servizi informativi (come quelli basati su regole predefinite), ma anche per generare risposte più flessibili e contestualmente adeguate, al fine di rendere le interazioni più naturali e coinvolgenti per gli utenti.

Il più noto e diffuso sistema di questo tipo è stato prodotto e reso accessibile al pubblico della rete Internet dalla società commerciale statunitense Open AI con il nome di *ChatGPT* (acronimo di Generative Pre-trained Transformer), sistema esperto di apprendimento automatico (*machine learning*), allenato su vaste quantità di testo e fondato su un ampio modello di linguaggio (*Large Language Model*) capace di produrre testo autonomo e sofisticato, con l'obiettivo di comprendere ed elaborare le complesse strutture semantiche della lingua.

Un elemento innovativo di questa tipologia di *chatbot* risiede nella modalità con cui i sistemi di intelligenza artificiale gestiscono il trattamento del linguaggio

⁷ Q. Yang, M. Wooldridge (ed.), *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, Buenos Aires, Argentina, 25-31 July 2015, AAAI Press, Palo Alto, California, USA, in <http://ijcai.org/papers15/contents.php>

⁸ P. Moro (a cura di), *Etica, diritto e tecnologia. Percorsi dell'informatica giuridica contemporanea*, FrancoAngeli, Milano, 2021.

⁹ P. Moro, "Libertà del robot? Sull'etica delle macchine intelligenti", in R. Brighi, S. Zullo (a cura di), *Filosofia del diritto e nuove tecnologie. Prospettive di ricerca tra teoria e pratica*, Atti del XXIV Congresso della Società Italiana di Filosofia del Diritto, Aracne, Roma, 2015, pp. 525-544.

naturale, includendo funzioni quali la risposta a interrogativi, la traduzione automatica, l'interpretazione dei testi e la loro sintesi. Queste capacità non sono sviluppate mediante un processo di apprendimento supervisionato basato su collezioni di dati specifiche per ciascuna attività. Piuttosto, questi sistemi acquisiscono autonomamente tali abilità, senza bisogno di direttive specifiche, attraverso un processo di formazione ininterrotto e graduale su volumi crescenti di informazioni, attingendo da milioni di pagine disponibili sulla rete Internet¹⁰.

Queste domande, che toccano anche il dibattito scientifico sull'interazione tra soggetto umano e persona digitale, possono essere esplorate attraverso un'analisi critica che oltrepassa la pura tecnologia dei sistemi cibernetici, attualmente definiti in modo generico "intelligenze artificiali", in base a principi di logica matematica e informatica.

Invece, si dovrebbe considerare la visione antropologica che giustifica questo modo di intendere la tecnologia. Molti ricercatori nelle discipline naturali, etiche e sociali continuano a confrontare l'essere umano con la macchina, spesso seguendo le teorie del post-umanesimo contemporaneo¹¹: tali concezioni promuovono una comprensione limitata e controversa della natura umana, secondo un'idea che affonda le radici nel razionalismo moderno e che è ancora prevalente nel pensiero scientifico attuale.

Infatti, attraverso l'uso di tecnologie avanzate e scienze sempre più raffinate, l'uomo moderno cerca da tempo di creare una sorta di "anima artificiale" nelle macchine, che non soltanto non possiedono un corpo biologico, ma ne sono sempre più spesso prive e sono gestite soltanto da *software*. Questo atteggiamento epistemologico indica una chiara direzione verso un obiettivo ben definito che la scienza cerca di raggiungere nei sistemi meccatronici, cercando di emulare le principali funzioni psichiche umane che, all'origine della filosofia occidentale, Platone suddivide nella nota tripartizione dell'anima tra ragione, volontà e passione¹².

Platone sostiene che l'equilibrio di queste tre facoltà spirituali riflette la giustizia nelle relazioni sociali e raccomanda di mantenere uniti questi aspetti dell'anima per comprendere appieno la natura umana.

In contrasto con questa visione olistica, per ragioni di opportunità politica o di profitto economico, l'era tecnologica e postmoderna si sta concentrando sull'amplificazione nelle macchine di intelligenza, volontà ed emozione in una sequenza esponenziale, seguendo un approccio individualistico e frammentato che riduce l'uomo ad un'unica dimensione¹³.

¹⁰ A. Radford, J. Wu, R. Child., D. Luan, D. Amodei, I. Sutskever, "Language Models are Unsupervised Multitask Learners", in *OpenAI blog*, 2019.

¹¹ R. Peppered, "The Posthuman Manifesto", in *Kritikos*, (2005), n. 2, pp. 1-16.

¹² Platone, *Repubblica*, 439 a-441 c, in *Opere*, vol. II, Laterza, Bari, 1967, pp. 262-265.

¹³ H. Marcuse, *One-Dimensional Man. Studies in the Ideology of Advanced Industrial Society*, Beacon Press, Boston, 1964.

2. Congegni intelligenti

L'intelligenza artificiale non è un fenomeno nuovo. La locuzione nasce dopo che alcuni matematici del secolo scorso considerano appropriato raffrontare la natura dei processi mentali di elaborazione e apprendimento delle informazioni alla struttura biologica dei processi naturali e sociali dell'attività fisiopsichica, al fine di realizzare il sogno di costruire una "macchina per pensare".

Nel 1936, Alan Turing progetta tale dispositivo, analizzando il processo razionale di calcolo e riferendosi agli "stati mentali" per definire le configurazioni interne della macchina e alla costruzione di un "cervello" per delineare l'esecuzione del suo progetto¹⁴.

Nel 1948, Norbert Wiener collega espressamente la ricerca sugli esseri viventi a quella sulle macchine per garantire un adeguato controllo su entrambi e sostiene l'impossibilità teorica di distinguere il sistema nervoso umano da un sistema di comunicazione meccanico come quello telefonico¹⁵.

In questo contesto storico e scientifico nasce la metafora "intelligenza artificiale".

Il 31 agosto 1955 negli Stati Uniti, durante gli studi della scienza dei calcolatori elettronici (*computer science*) e della cibernetica (*cybernetics*), lo scienziato nordamericano John McCarthy propone ad alcuni colleghi (come Marvin Minsky, Nathaniel Rochester e Claude Shannon) di designare con il sintagma *artificial intelligence* il titolo di un progetto estivo di ricerca relativo ad un modello computerizzato della mente umana, che sviluppava i presupposti della cibernetica di Wiener¹⁶.

Nel *paper proposal* del 1955 sono già presenti i tratti essenziali dell'intelligenza artificiale contemporanea: automazione e simulazione delle funzioni del cervello umano; utilizzo del linguaggio naturale; creazione di reti neurali; teoria della dimensione del calcolo; auto-miglioramento delle macchine; astrazioni dai dati sensoriali; casualità e creatività.

Come afferma successivamente Pat Hayes nel suo *naive physics manifesto*¹⁷, ciò che si vuole rappresentare con le applicazioni dell'intelligenza artificiale è il modo in cui le persone vedono il mondo, non il mondo come viene visto dalla fisica, con la conseguente necessità di elaborare modelli che riguardano non solo oggetti od eventi, ma anche percezioni e azioni inerenti alle modalità di comprensione del mondo da parte delle persone.

¹⁴ A.M. Turing, "On computable numbers with an application to the Entscheidungsproblem", in *Proceedings of the London Mathematical Society*, 42 (1936), pp. 230-265.

¹⁵ N. Wiener, *Cybernetics, or control and communication in the animal and the machine*, Hermann, J. Wiley, New York, Paris, 1948.

¹⁶ J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955", in *AI Magazine*, 27 (2006), n. 4, p. 12.

¹⁷ P.J. Hayes, "The naive physics manifesto", in *Expert Systems in the Micro-Electronic Age*, Edinburgh University Press, Edinburgh, 1978.

Oggi giorno, questa comparazione scientifica tra intelligenza artificiale e ragione umana presuppone l'adesione alla tesi dell'intelligenza artificiale "forte" ed appare inevitabilmente fondata sull'assioma concettuale della cibernetica, secondo cui la macchina per pensare emula la ragione che calcola. Infatti, lo sforzo di innalzare il dispositivo tecnologico dotato di intelligenza artificiale al livello delle facoltà umane si realizza attraverso operazioni di formalizzazione che, in sostanza, si traducono nella costruzione di algoritmi informatici che governano ed accelerano il presunto ragionamento e la conseguente azione del sistema di intelligenza artificiale.

Il procedimento calcolante custodisce la sua base computante anche nei sistemi informatizzati, che costituiscono elaboratori evoluti della tecnica combinatoria sviluppata dai logici più importanti dell'epoca barocca, come Pascal e Leibniz, per connettere aritmetica e meccanica e fondare così nell'aurora della scienza moderna un'autentica *mathesis universalis*¹⁸.

Questa concezione si mostra nell'alveo teoretico del razionalismo, secondo il quale l'uomo sarebbe capace di conoscere ciò che è bene e ciò che è male attraverso l'uso della propria ragione, e ha contribuito a fondare i programmi di intelligenza artificiale sul metodo scientifico¹⁹, riconducendo a schemi comuni di origine algoritmica le macchine e i sistemi biologici capaci di esibire un comportamento finalizzato, confermando il presupposto matematico della cibernetica²⁰.

Tuttavia, la rappresentazione informatizzata della conoscenza si sta realizzando in determinati sistemi di intelligenza artificiale soprattutto con l'apprendimento automatico (*machine learning*), favorito dagli studi di statistica e delle neuroscienze, perché basato su molteplici esempi (denominati *big data*) tratti dal grande archivio di Internet e sull'elaborazione di sistemi di intelligenza artificiale (denominati reti neurali) simili al cervello umano.

Questo approccio statistico, basato sui dati elaborati con le tecnologie di *machine learning* e di ottimizzazione stocastica, ha sostituito la logica deduttiva, trasformando l'inferenza intelligente nel modello inverso, risolto dalla massimizzazione induttiva di una certa quantità probabilistica²¹.

Intelligenza artificiale e rete Internet sono così diventate le più importanti tecnologie convergenti dell'epoca contemporanea, aumentando il loro predominio in via esponenziale, perché interagiscono tra loro e si implicano reciprocamente, pur sempre in dipendenza di algoritmi di apprendimento automatico: questa soggezione permane sia nel caso di algoritmo supervisionato, ossia quando il

¹⁸ W. Schmidt-Biggemann, *Topica universalis. Eine Modellgeschichte humanistischer und barocker Wissenschaft*, Meiner, Hamburg, 1983, pp. 156 ss.

¹⁹ D.A. Gillies, *Artificial intelligence and scientific method*, Oxford University Press, Oxford, 1996, trad. it., *Intelligenza artificiale e metodo scientifico*, Cortina, Milano, 1998.

²⁰ G. Tamburrini, *I matematici e le macchine intelligenti. Spiegazione e unificazione nella scienza cognitiva*, Mondadori, Milano, 2002.

²¹ N. Cristianini, "On the current paradigm in artificial intelligence", in *AI Communications* 27 (2011), pp. 37-43.

sistema (per esempio, la rete neurale) apprende il nesso che unisce dati particolari e ipotesi generali, imparando a classificare esempi simili, sia nel caso di algoritmo per rinforzo, ossia quando il sistema interagisce con l'ambiente ed esegue gli esempi di comportamenti più vicini e simili al risultato, che quindi migliorano l'azione da eseguire, evitando quelli più distanti e diversi, che sono considerati errati.

3. Congegni morali

I sistemi dotati di intelligenza artificiale non sono più soltanto macchine calcolanti o *robot* che sostituiscono parti ripetitive del lavoro umano, come indica l'etimologia del termine, che deriva dal ceco *robota* e che è stato introdotto dallo scrittore Karel Čapek nel dramma fantascientifico *R.U.R.*²².

Infatti, i dispositivi più sofisticati ed evoluti governati dall'intelligenza artificiale appaiono anche entità agenti o congegni morali, in grado di deliberare e di assumere decisioni che, osservate dall'esterno, appaiono simili a manifestazioni di volontà soggettiva, soggette a regole di comportamento elaborate dalla roboetica e fondate su valori scientifici, culturali e tecnici universalmente condivise e dirette alla tutela del benessere della persona umana²³.

È difficile ammettere l'attribuzione di diritti e doveri alle macchine intelligenti, sebbene queste operino con sempre maggiore autonomia in analogia con il modo di pensare e di agire dell'uomo, tanto che (come documenta l'approvazione in Europa del nuovo AI Act), appare ormai in corso di evoluzione la regolazione della specifica responsabilità dei sistemi cibernetici, richiesta dalla discussione avviata negli ultimi anni nella robotica e nella bionica²⁴ e considerata urgente ed indifferibile nella tecnoetica, soprattutto al fine di preservare la tutela e l'integrità delle persone attraverso l'imposizione di limiti etici e giuridici alle applicazioni della ricerca tecnologica²⁵.

Peraltro, il *robot* che interagisce con l'uomo in modo autosufficiente deve affrontare una serie di dilemmi morali che impongono alla macchina intelligente la necessità di deliberare una scelta fra più alternative, come, per esempio, costringere un anziano ad assumere un farmaco.

Introducendo il concetto di "macchine morali" per designare questi sistemi particolarmente sviluppati, è stato osservato che l'implementazione di regole morali

²² K. Čapek, *Rossumovi univerzální robot*, Praha, 1920, trad. it., *R.U.R. Rossum's Universal Robots*, a cura di A. Catalano, Venezia, Marsilio, 2015.

²³ G. Veruggio, "The birth of roboethics", in *Institute of Electrical and Electronics Engineers International Conference on Robotics and Automation*, Workshop on Roboethics, Barcellona, 2005.

²⁴ A. Montanari, "Questioni di tecnoetica in intelligenza artificiale, robotica e bionica", in P. Moro (a cura di), *Etica informatica diritto*, FrancoAngeli, Milano, 2008, pp. 33-50.

²⁵ A. Fabris, G. Tamburrini, "Robot al di qua del bene e del male", in *Vita e Pensiero*, (2006), n. 3, pp. 85-93.

o giuridiche che permettono al *robot* di prendere decisioni consapevoli e comportarsi in modo giusto oppure sbagliato può essere programmata con algoritmi di istruzione ma anche di apprendimento dalla condotta che l'automa pone in essere, ma si è precisato giustamente che tale questione supera i confini di una forma controllabile di intelligenza artificiale²⁶.

Inoltre, allo stato attuale, la possibilità di stabilire un'analogia tra stato mentale e logica della macchina avviene solo sul piano dell'accettabilità dei risultati e non della identità delle procedure di ragionamento e di decisione: tali metodologie sono propriamente umane e sono radicalmente differenti nel calcolatore che, tuttavia, può arrivare ai medesimi esiti dimostrativi²⁷.

In ogni caso, se il *robot* diventa sempre più autonomo nell'esecuzione dei propri compiti a causa dell'accelerazione dei processi autonomi di *machine learning*, come può accadere ad una macchina cibernetica impiegata per scopi militari oppure per attività di polizia, esso può diventare centro d'imputazione, del tutto analogo all'uomo, di atti responsabili per gli eventuali danni provocati dal loro comportamento, con la conseguenza che gli agenti intelligenti (che sono composti di *hardware* e di *software*) potrebbero diventare titolari di diritti e doveri simili a quelli delle persone e, dunque, entità penalmente imputabili.

Invero, è evidente la problematicità semantica del linguaggio che si utilizza per tentare di spiegare questa dimensione "soggettiva" dell'intelligenza artificiale con categorie antropologiche ormai cristallizzate nella storia e nel pensiero occidentale, quali quelle di autonomia, volontà, responsabilità: si tratta di un lessico del tutto insufficiente ad illustrare adeguatamente una presunta libertà di agire di una entità meccanica. La responsabilità delle macchine intelligenti deve essere ripensata con le nuove categorie del tecnodiritto, come è già stato notato nella dottrina italiana maggiormente attenta al fenomeno, sia nel settore della responsabilità civile²⁸ che in quello della responsabilità penale²⁹.

La costruzione del cosiddetto *imitation game* da parte di Alan Turing, secondo il quale un sistema artificiale può essere definito intelligente se riesce a convincere una persona che il suo comportamento intellettuale è quello di un essere umano medio, è indubbiamente condizionata da un'interpretazione della coscienza e della volontà della macchina basata sugli effetti della condotta reattiva agli impulsi esterni: tale interpretazione deriva tuttora dalla psicologia comportamentale di John Watson, secondo cui la condotta esplicita sarebbe l'unica unità di analisi

²⁶ W. Wallach, C. Allen, *Moral machines. Teaching robots right from wrong*, Oxford University Press, Oxford, 2009.

²⁷ S. Crafa, "Artificial Intelligence and Human Dialogue", in *Journal of Ethics and Legal Technologies*, (2019), n. 1, pp. 44-56.

²⁸ U. Ruffolo, "Per i fondamenti di un diritto della robotica self-learning, dalla machinery produttiva all'auto driverless: verso una responsabilità da algoritmo?", in Id., *Intelligenza Artificiale e Responsabilità*, Giuffrè, Milano, 2018, pp. 1-30.

²⁹ S. Riondato, "Robot: talune implicazioni di diritto penale", in P. Moro, C. Sarra (a cura di), *op. cit.*, pp. 85-98.

scientificamente studiabile, in quanto direttamente osservabile, mentre l'introspezione non potrebbe fornire alcun dato affidabile³⁰.

La costruzione antropomorfa della scienza informatica odierna, che assegna alle operazioni effettuate da sistemi informatici o cibernetici la nozione di volontà come reazione della macchina (*output*) ad un impulso dell'utente (*input*), condiziona l'approccio riduzionista della "società della mente" di Marvin Minsky, secondo il quale il cervello è una macchina altamente complessa paragonabile ad una società organizzata da una rete di interconnessioni gerarchiche e competitive tra una molteplicità di componenti fortemente diversi fra loro, denominati appunto "agenti della mente"³¹.

Occorre superare un'interpretazione semplificata e positivista dell'essere umano, riconoscendo che la volontà umana che il *robot* riproduce non è una decisione di agire basata sulla reazione ad impulsi esterni, ma una tensione verso un fine indisponibile che indica una qualità sostanziale dell'uomo nel suo intero. Dunque, prima di implementare la volontà umana nella macchina e sempre che tale operazione possa essere ammissibile e anche possibile, sarebbe necessario definire più compiutamente che cosa la volontà sia effettivamente, senza limitare l'indagine dell'essenza del soggetto all'analisi biologica delle sue facoltà psicologiche, come sembra accadere nella dominante prospettiva neuroscientifica.

Il pensiero e l'azione, così come l'intelligenza e la volontà, sono forme di vita, ma tali forme sono originarie e trovano il proprio fine nel loro stesso manifestarsi, sicché dimostrano di possedere una struttura anche soltanto biologica profondamente diversa da quelle che sembrano oggi ricostruibili con i sistemi cibernetici, come si desume dal sistema metabolizzante del corpo vivente, nel quale il metodo di produzione e di scambio di energia è anche il risultato dell'attività metabolica³².

4. Congegni emotivi

In un terzo aspetto emulativo delle facoltà psichiche umane, gli studi e le applicazioni dell'intelligenza artificiale hanno già da tempo cercato di progettare e realizzare congegni emotivi, governati da software in grado di riconoscere ed esprimere emozioni (*affective computing*)³³.

L'interazione umana con i sistemi di intelligenza artificiale sta portando allo sviluppo di macchine affettive (*sentiment machines*) che valutano la reazione emotiva dell'utente e interagiscono con il medesimo, sollevando alcune

³⁰ J.B. Watson, "Psychology as a behaviorist view", in *Psychological review*, 20 (1913), n. 2, pp. 158-177.

³¹ M.L. Minsky, *The Society of Mind*, Simon and Schuster, New York, 1986.

³² H. Jonas, "Ist Gott ein Mathematiker? Vom Sinn des Stoffwechsels", in *Organismus und Freiheit. Ansätze zu einer philosophischen Biologie*, Vandenhoeck & Ruprecht, Göttingen, 1973.

³³ R.W. Picard, *Affective computing*, MIT Press, Cambridge (MA), 1997.

problematiche etiche e giuridiche in materia di riservatezza e di autodeterminazione del singolo.

Le *sentiment machines* rappresentano l'avanguardia tecnologica dell'intelligenza artificiale, che esplora le intersezioni tra macchine calcolanti ed emozioni umane, e sono il risultato di una ricerca che mira a sviluppare sistemi cibernetici capaci di riconoscere, interpretare e, in alcuni casi, replicare le emozioni umane.

Un aspetto fondamentale degli studi sui computer affettivi è la capacità delle macchine di percepire e reagire alle emozioni umane in modo appropriato, con l'esigenza di realizzare avanzamenti tecnici in termini di riconoscimento facciale o elaborazione del linguaggio naturale ma anche con la necessità di affrontare sfide etiche e filosofiche su natura e limiti degli interventi sulle esperienze emotive umane.

Per esempio, nelle applicazioni di marketing finanziario e di comunicazione politica, ha assunto peculiare importanza l'analisi computazionale delle emozioni (*sentiment analysis*) basata sulla ricerca ed esame delle opinioni (*opinion mining*) pubblicate sulla rete Internet e reperibili principalmente sui maggiori social network³⁴.

Inoltre, la percezione emotiva degli esseri umani è studiata nel campo della robotica sociale e dell'assistenza sanitaria per programmare "robot di compagnia" in modo che possano simulare le emozioni nelle loro interazioni con gli esseri umani, con evidenti conseguenze non solo etiche e sociali, ma anche giuridiche in materia di dignità umana e diritti fondamentali³⁵.

Gli esempi dimostrano la rilevanza di "macchine emotive" particolarmente progredite e l'impatto giuridico che potrà avere su tali sistemi il nuovo AI Act nel territorio dell'Unione Europea. Infatti, basti rammentare che l'articolo 5 del nuovo regolamento europeo considera vietate pratiche di manipolazione del comportamento attraverso tecniche subliminali o di sfruttamento delle vulnerabilità di minori o disabili o di profilazione illegittima; attribuzione di un punteggio sociale basato sull'AI con trattamento sfavorevole per le persone da parte di autorità pubbliche; ricorso a sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto, salvo talune eccezioni.

La capacità degli attuali sistemi di intelligenza artificiale di esprimere emozioni si fonda sull'abilità di imitare le espressioni del volto umano attraverso algoritmi di apprendimento automatico e tecniche di *deep learning*. Questi "computer affettivi" possono analizzare le espressioni facciali elaborando rapidamente miliardi di immagini, grazie all'accesso a enormi quantità di dati provenienti da diverse fonti, e si affiancano a modelli computazionali avanzati che

³⁴ B. Liu, "Sentiment analysis and opinion mining", in *Synthesis Lectures on Human Language Technologies*, 5 (2012) n. 1, pp. 1-167.

³⁵ L. Devillers, "Human-Robot Interactions and Affective Computing: The Ethical Implications", in J.S. von Braun, M. Archer, G.M. Reichberg, M. Sánchez Sorondo (eds.), *Robotics, AI, and Humanity*, Springer, Cham, 2021, pp. 205-211.

imitano aspetti dell'intelligenza emotiva attraverso la creazione di reti neurali artificiali che replicano alcuni aspetti delle emozioni, come se fossero capaci di "sentire".

Tuttavia, nonostante questi progressi, è noto che ogni sistema di intelligenza artificiale resta confinato ad elaborare ed eseguire compiti specifici, come la combinazione di parole e il riconoscimento di immagini o suoni, ma non riesce ancora a simulare una forma di intelligenza generale che richiederebbe un'interazione complessa di funzioni cognitive coordinate e dirette alla gestione di più azioni nello stesso momento.

Infatti, le ricerche neuroscientifiche documentano che anche alle macchine affettive manca il senso dell'intero, come dimostrano gli studi sulle relazioni tra intelligenza artificiale e le dinamiche *whole brain*, che riguardano l'interazione tra aree dell'intero cervello che non sono indipendenti ma che comunicano tra loro e partecipano a un sistema di relazioni complesso e articolato³⁶.

In ogni caso, anche le macchine emotive fondano la propria struttura tecnologica sulle acquisizioni dell'intelligenza artificiale ed appaiono influenzate dalla visione critica della natura umana che le neuroscienze hanno proposto nel dibattito degli ultimi anni, opponendosi alla tradizionale concezione dualistica del soggetto, di origine cartesiana, che contrappone le facoltà intellettive alla struttura della materia e che appiattisce il soggetto umano nel campo dei fenomeni³⁷.

5. Congegni inconsapevoli

Nel presente lavoro si è notato che la crescita esponenziale degli studi e delle applicazioni dell'intelligenza artificiale sta percorrendo un itinerario emulativo delle facoltà umane di ragionare, di decidere e di emozionarsi.

Questa imitazione delle principali qualità psichiche da parte delle macchine, guidata da scienziati e programmatori in una sequenza quasi pedagogica, sembra l'esito narcisistico di una prospettiva antropocentrica e riduzionistica del soggetto umano, che ricompona le proprie funzioni con la fabbricazione di frammenti digitali del soggetto.

È il risultato della dogmatica e mai adeguatamente superata rappresentazione meccanica dell'intelligenza artificiale, frutto del positivismo cibernetico del secolo scorso, con la conseguenza che l'evoluzione dei *robot* continua ad essere un processo di sviluppo di alcune facoltà soggettive che sono ritenute univoche, come la ragione calcolante, la volontà reattiva o l'emozione oggettivata, ma che non costituiscono gli autentici e compiuti modi di definire queste caratteristiche della natura umana.

³⁶ H. Mizutani, M. Ueno, N. Arakawa, H. Yamakawa, "Whole brain connectomic architecture to develop general artificial intelligence", in *Procedia Computer Science*, 123 (2018), pp. 308-313.

³⁷ S. Fuselli, *Diritto, neuroscienze, filosofia. Un itinerario*, FrancoAngeli, Milano, 2014.

In effetti, la coscienza non è riducibile ad una sola delle sue funzioni e, dunque, il tentativo di paragonare la mente umana all'intelligenza artificiale è destinato a fallire: la coscienza è una proprietà dell'intera persona e deve essere considerata nella totalità che contraddistingue l'intelligenza umana, che non può essere riprodotta, quantomeno nella sua completezza, attraverso modelli computazionali anche molto evoluti, giacché essa genera un problema che non è computabile³⁸.

Forse l'intelligenza artificiale è un modo di pensare, ma resta una forma della ragione calcolante essenzialmente ridotta al codice binario che governa tuttora la scienza informatica. Invece, l'uomo è dotato di una mente colorata, come ricorda Omero a proposito del multiforme ingegno di Ulisse³⁹, e vive di plurime dimensioni della soggettività che, come ha accertato la psicanalisi del Novecento, provengono anche dall'inconscio: come osserva Freud, la vita psichica si svolge in modalità performative anche nelle sue forme secondarie – *lapsus*, dimenticanze, sogni – che indicano un'organizzazione alternativa della razionalità⁴⁰ e che sono del tutto ignote all'intelligenza meccanica che, quando sbaglia, esegue un'operazione che è giudicata falsa in base all'algebra di Boole e, dunque, non commette affatto un *lapsus*.

Dunque, l'equivalenza tra mente umana e cervello elettronico è inammissibile, perché le macchine calcolanti sono inconsapevoli. Infatti, applicando gli algoritmi di programmazione che ogni *computer* deve comunque eseguire, il dispositivo cibernetico elabora un ragionamento meccanico senza comprendere ciò che dice o ciò che fa.

Tutte le logiche di tipo formale, che sono largamente utilizzate nelle macchine intelligenti, si manifestano tramite un lessico e una sintassi di tipo rigidamente simbolico, analizzabile seguendo regole di inferenza che hanno natura algoritmica, deduttiva o induttiva. La comprensione del linguaggio e l'interpretazione delle regole inferenziali da parte dell'operatore che ne fa uso è indipendente dallo sviluppo e dall'esito del procedimento inferenziale.

Nel suo famoso esperimento mentale, noto come “stanza cinese” ed illustrato per criticare i sostenitori dell'intelligenza artificiale “forte”, John Searle ha contestato che la mente possa essere un programma, perché il *computer* non capisce ciò che sta facendo quando elabora un'operazione mentale attraverso il linguaggio simbolico, essendovi una distanza incolmabile tra sintassi e semantica, ossia tra struttura grammaticale del linguaggio e comprensione significante delle operazioni combinatorie⁴¹.

³⁸ F. Faggin, *Irriducibile. La coscienza, la vita, i computer e la nostra natura*, Mondadori, Milano, 2022.

³⁹ P. Citati, *La mente colorata. Ulisse e l'Odissea*, Mondadori, Milano, 2002.

⁴⁰ S. Freud, *Vorlesungen zur Einführung in die Psychoanalyse*, trad. it., *Introduzione alla psicoanalisi* [1915], in *Opere di Sigmund Freud*, VIII, Bollati Boringhieri, Torino, 2000.

⁴¹ J.R. Searle, “Minds, brains, and programs”, in *Behavioral and Brain Sciences*, (1980), n. 3, pp. 417-457.

Le argomentazioni di Searle colgono il punto nevralgico di ogni teoria sull'intelligenza artificiale e, pur essendo tuttora oggetto di dibattito scientifico⁴², consentono indubbiamente di escludere che la logica proposizionale booleana, spesso considerata un postulato del ragionamento informatizzato, possa realizzare un qualsiasi sistema complesso in grado di superare la morfologia sintattica del linguaggio per giungere alla sua piena comprensione semantica.

Nel dominio dell'informatica, l'incapacità di un sistema di intelligenza artificiale di riconoscere e superare i propri limiti è provata dalla presenza di problemi non computabili (o indecidibili) che non sono risolvibili dall'elaboratore elettronico. Tale scenario si manifesta nell'impossibilità di sviluppare un algoritmo di verifica per programmi che possa determinare, data una serie specifica di premesse, se un dato software terminerà il suo processo con un risultato logicamente coerente o se, al contrario, continuerà all'infinito a produrre un esito incoerente⁴³.

Pertanto, l'approccio computazionale, che rappresenta la base teorica per la ricerca e lo sviluppo nel campo dell'intelligenza artificiale, deve affrontare la specifica capacità del pensiero umano di autotrascendersi e di ridefinirsi, che rappresenta una delle caratteristiche fondamentali e irriducibili dell'intelligenza umana: questa facoltà soggettiva è inesauribile e imprevedibile, sicché appare difficilmente afferrabile dalla computazione⁴⁴.

6. Persone elettroniche

La possibilità di attribuire una vera e propria "soggettività robotica" ad una macchina particolarmente sofisticata, che presenta forma materiale (*hardware*) o immateriale (*software*) e che agisce anche senza controllo umano, sta attraendo su di sé nozioni e metafore delle scienze matematiche ed informatiche del tutto differenti tra loro (algoritmo, intelligenza artificiale, *chat bot*) in base ad una visione individualista, che è tipica dell'epoca contemporanea e che sta influenzando non solo la cultura e la società, ma anche la più tipiche categorie dell'etica e del diritto dell'era moderna, come la responsabilità.

Poiché le nozioni di uomo e di macchina sono incommensurabili, è opportuno evitare di applicare ad un software cibernetico particolarmente evoluto l'idea antropocentrica della responsabilità soggettiva, talora presupposta nel pensiero

⁴² S. Harnad, "What's Wrong and Right About Searle's Chinese Room Argument", in M. Bishop, J. Preston (eds.), *Essays on Searle's chinese room argument*, Oxford University Press, Oxford, 2011.

⁴³ D. Harel, *Computers Ltd.: What They Really Can't Do*, Oxford University Press, Oxford and New York, 2000.

⁴⁴ F. Chiereghin, *L'eco della caverna. Ricerche di filosofia della logica e della mente*, Il Poligrafo, Padova, 2004, pp. 200-201.

tecnologico della postmodernità, valutando attentamente la rilevanza giuridica del passaggio di un'entità robotica da *res* a *persona*⁴⁵.

Tuttavia, la macchina dotata di intelligenza artificiale avanzata opera ormai attivamente nell'odierna "infosfera", nella quale il tutto viene prima del soggetto agente e l'uomo non è l'unico soggetto cui attribuire una dignità morale, diventando così vero e proprio "organismo informazionale" (*infor*g), le cui condotte, intenzionali o meno, possono produrre effetti sulle entità interconnesse⁴⁶.

Soprattutto nel suo approccio odierno, il sistema di intelligenza artificiale ottiene un'acquisizione di cognizioni e condotte che restano indipendenti dalla programmazione informatica e, quindi, dall'intervento umano⁴⁷, e costruisce automaticamente un modello dall'analisi dei dati su cui è addestrato, generando valutazioni e previsioni sui nuovi casi ad esso sottoposti in base al continuo aggiornamento dei dati medesimi, migliorando il modello predittivo attraverso l'ottimizzazione stocastica e la massimizzazione induttiva di una certa quantità probabilistica⁴⁸.

È indubbio che l'apprendimento automatico delle macchine può forse puntare a raggiungere l'autentico "algoritmo dominante"⁴⁹ e che gli esiti del modello creato dall'intelligenza artificiale e le ragioni che ne giustificano le decisioni non possono essere spiegati dal codice sorgente o dai risultati osservabili della procedura impostata, giacché il funzionamento interno del sistema resta oscuro agli stessi programmatori e può essere considerato una *black box*⁵⁰.

Infatti, tenendo presente il progressivo aumento di autosufficienza di sistemi sempre più prossimi a compiere attività sempre meno condizionata, l'articolo 3 numero 1) del nuovo AI Act dell'Unione Europea designa il "sistema di intelligenza artificiale" come "un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali".

Nel *considerandum* 12 del regolamento, si precisa che "la capacità inferenziale di un sistema di IA trascende l'elaborazione di base dei dati e consente l'apprendimento, il ragionamento o la modellizzazione" e che "l'adattabilità che un

⁴⁵ U. Ruffolo, "Il problema della 'personalità elettronica'", in *Journal of Ethics and Legal Technologies*, (2020), n. 1, pp. 75-88.

⁴⁶ L. Floridi, *La quarta Rivoluzione*, Raffaello Cortina, Milano, 2017.

⁴⁷ S. Russell, P. Norvig, *Artificial Intelligence. A Modern Approach*, Prentice Hall, Englewood Cliffs (N.J.), 2010.

⁴⁸ N. Cristianini, "On the current paradigm in artificial intelligence", in *AI Communications*, 27 (2011), pp. 37-43.

⁴⁹ P. Domingos, *The Master Algorithm. How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, New York, 2015.

⁵⁰ F. Pasquale, *The black box society: The secret algorithms that control money and information*, Harvard University Press, Cambridge (MA), 2015.

sistema di IA potrebbe presentare dopo la diffusione si riferisce alle capacità di autoapprendimento, che consentono al sistema di cambiare durante l'uso”.

Peraltro, occorre considerare che la soggettività umana si distingue dal sistema di intelligenza artificiale, che può animare un dispositivo immateriale oppure un corpo sintetico, anche e soprattutto per la condizione biologica, che caratterizza specificamente l'esistenza umana⁵¹.

Inoltre, il tratto originario e irripetibile della soggettività umana, che la distingue dall'intelligenza meccanica, è la relazione con l'altro da sé⁵², la quale costituisce la finalità e la metodologia del diritto, che è un'attività di tutela e promozione della dignità umana⁵³. Anzi, è stato giustamente osservato che, nell'ipercomunicazione dell'era tecnologica, la scomparsa del senso dell'Altro è il motivo ontologico per cui l'utilizzo continuo del dispositivo digitale accresce il vuoto e aggrava la solitudine dell'umanità contemporanea⁵⁴.

Tuttavia, è indubbio che sono sempre più rilevanti le interazioni sociali ed autonome degli agenti software dotati di intelligenza artificiale⁵⁵ e che la capacità di apprendimento del robot evoluto è interattiva (*interactive learning*) e incrementa progressivamente la propria autonomia non soltanto apprendendo dall'ambiente e dagli oggetti che incontra, ma anche cooperando con l'essere umano, come avviene nella robotica collaborativa⁵⁶. L'impatto della regolazione algoritmica della società (per esempio, attraverso attività di *crowdsourcing* o di gestione della reputazione) ha anche suggerito la nozione di “macchina sociale”, che genera evidenti conseguenze etiche e giuridiche per la tutela dell'autonomia individuale e dell'ordine pubblico⁵⁷.

Pertanto, non è più azzardato ipotizzare che la macchina intelligente, presuntivamente dotata di facoltà artificiali analoghe, se non superiori, a quelle umane, possa essere titolare di una specie di autodeterminazione individuale e, quindi, di una sorta di libero arbitrio, come ha sostenuto John McCarthy⁵⁸. Infatti, se tale qualificazione fosse accettabile, occorrerebbe rivedere l'idea antropocentrica dell'intelligenza artificiale e prendere atto che l'agente robotico particolarmente

⁵¹ P. Benanti, *The Cyborg: corpo e corporeità nell'epoca del post-umano. Prospettive antropologiche e riflessioni etiche per un discernimento morale*, Cittadella Editrice, Assisi, 2012.

⁵² G. De Bona, “Verso la tutela giuridica dei sistemi intelligenti. Prospettive critiche della soggettività robotica”, in *Journal of Ethics and Legal Technologies*, 4 (2022) n. 2, pp. 51-104.

⁵³ A.C. Amato Mangiameli, “IA e diritto. In luogo di una introduzione”, in *Journal of Ethics and Legal Technologies*, 5 (2023), n. 2, pp. 81-94.

⁵⁴ Byung-Chul Han, *Undinge: Umbrüche der Lebenswelt*, Ullstein Verlag, Berlin, 2021, trad.it., *Le non cose. Come abbiamo smesso di vivere il reale*, Einaudi, Torino, 2022, p. 39.

⁵⁵ G. Teubner, “Die anonyme Matrix: Zu Menschenrechtsverletzungen durch ‚private‘ transnationale Akteur”, in *Der Staat*, 45 (2006), n. 2, pp. 161-187.

⁵⁶ M.A. Goodrich, A.C. Schultz, “Human-Robot Interaction: A Survey”, in *Foundations and Trends in Human-Computer Interaction*, (2007), n. 3, pp. 203-275.

⁵⁷ N. Cristianini and T. Scantamburlo, “On social machines for algorithmic regulation”, in *AI & Society*, Springer, Verlag, 35 (2020), pp. 645-662.

⁵⁸ J. McCarthy “Free Will-even for Robots”, in *Journal of Experimental and Theoretical Artificial Intelligence*, 12 (2000), n. 3, pp. 341-352.

s sofisticato potrebbe acquisire una certa forma di libertà soggettiva e, quindi, una pur imprecisata titolarità di diritti e di doveri⁵⁹ oppure di una certa capacità di agire, sempre che si ammetta l'applicabilità di tali categorie ad una macchina⁶⁰.

È da notare che la programmazione di decisioni consapevoli e l'applicazione di regole condivise può essere imposta al sistema robotico da algoritmi di istruzione e di apprendimento, ma tale questione supera i confini di una forma controllabile di intelligenza artificiale⁶¹. Difatti, le procedure formali di apprendimento automatico restano inferenze ipotetiche e non possono interpretare o risolvere problemi complessi della vita reale, come i conflitti di valore, che manifestano costante variabilità ed incertezza e che sono difficili da codificare attraverso misurazioni o funzioni numeriche⁶².

In definitiva, quanto sopra esaminato suggerisce di riflettere senza pregiudizi sulla questione della soggettività robotica, che non è paragonabile alla soggettività umana, ma che potrebbe essere attribuita a sistemi evoluti e complessi di intelligenza artificiale, le cui limitazioni e tutele dovrebbero essere indicate non dalle categorie giuridiche tradizionali, ma dalla costituzione di una nuova finzione, utilizzando il linguaggio visivo della retorica.

In sostanza, per designare un sistema robotico di intelligenza artificiale che opera in modo simile a quello umano, pare del tutto appropriato introdurre nel tecnodiritto contemporaneo una metafora simile a quella della persona giuridica (in inglese, *legal entity*), ben nota al linguaggio dei giuristi e inventata per risolvere problemi fondamentali del traffico commerciale e della responsabilità risarcitoria⁶³.

Infatti, ci si deve domandare se, come ha fatto per il momento il nuovo AI Act, potrà davvero essere trascurata dal legislatore europeo la proposta, contenuta nella Risoluzione del Parlamento europeo del 16 febbraio 2017, di costituire per i robot evoluti una innovativa finzione giuridica, assegnando la condizione di "persone elettroniche responsabili" a determinati sistemi di intelligenza artificiale.

La citata Risoluzione, recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica e basata sulla logica della sicurezza del prodotto commerciale e sulla gestione del rischio⁶⁴, contiene un esplicito invito al massimo organo legislativo europeo, e indirettamente gli Stati membri dell'Unione, "a esplorare, esaminare e valutare, nell'ambito della valutazione

⁵⁹ P. Moro, "Libertà del robot? Sull'etica delle macchine intelligenti", cit., pp. 525-544.

⁶⁰ E. Morotti, "Una soggettività a geometrie variabili per lo statuto giuridico dei robot", in F. Bilotta, F. Raimondi (a cura di), *Il soggetto di diritto. Storia ed evoluzione di un concetto nel diritto privato*, Jovene Editore, Napoli, 2020, pp. 291-306.

⁶¹ W. Wallach, C. Allen, *Moral machines. Teaching robots right from wrong*, Oxford University Press, Oxford, 2009.

⁶² Z. Lipton, "The mythos of model interpretability", in *ICML Workshop on Human Interpretability*, 2016, arXiv:1511.03677.

⁶³ F. Galgano, *Le insidie del linguaggio giuridico. Saggio sulle metafore nel diritto*, il Mulino, Bologna, 2010.

⁶⁴ N. Busto, "La personalità elettronica dei robot: logiche di gestione del rischio tra trasparenza e fiducia", in *Cyberspazio e diritto*, 18 (2017) n. 3, pp. 499-511.

d’impatto del suo futuro strumento legislativo, le implicazioni di tutte le soluzioni giuridiche possibili, tra cui: (...) l’istituzione di uno status giuridico specifico per i robot nel lungo termine, di modo che almeno i robot autonomi più sofisticati possano essere considerati come persone elettroniche responsabili di risarcire qualsiasi danno da loro causato, nonché eventualmente il riconoscimento della personalità elettronica dei robot che prendono decisioni autonome o che interagiscono in modo indipendente con terzi”.

Il documento legislativo del Parlamento Europeo, cui ha dato origine il percorso che ha portato all’approvazione del nuovo regolamento sull’intelligenza artificiale, è di evidente interesse per individuare il principio giuridico, peraltro tratto da una citazione letteraria, che risolva le probabili controversie tra uomini e robot: nelle premesse giustificative, infatti, la Risoluzione considera formalmente le quattro leggi di Asimov, tra le quali la fondamentale “legge zero” (“Un robot non può recar danno all’umanità e non può permettere che, a causa di un suo mancato intervento, l’umanità riceva danno”)⁶⁵, “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”.

Nel silenzio della legislazione, la costruzione di una nuova *fiction iuris* denominata “persona elettronica” oppure “entità elettronica” (in inglese *electronic person* oppure *electronic entity*) per designare la soggettività robotica potrebbe essere demandata alla giurisprudenza per consentire di risolvere alcune rilevanti questioni pratiche, come i conflitti provocati dall’interazione sociale tra sistemi di intelligenza artificiale e soggetti umani.

La natura e i limiti di questa finzione, che sta aprendo nuovi ed importanti orizzonti di ricerca dell’informatica giuridica contemporanea, dovranno essere certamente soppesate da riflessioni adeguate ai principi di ogni ordinamento sociale⁶⁶, ma potranno essere definite anche dal progresso tecnologico, la cui evoluzione porterebbe a forme graduali di riconoscimento della personalità elettronica, in base allo stadio di autonomia dei sistemi di intelligenza artificiale⁶⁷.

In ogni ipotesi, la finzione giuridica della “persona elettronica” non è affatto peregrina e appare del tutto compatibile con la dottrina secondo cui le categorie concettuali, che tradizionalmente identificano nell’ordinamento sociale la tutela e la responsabilità del soggetto umano, siano applicabili già ai calcolatori elettronici degli anni Ottanta del Novecento⁶⁸ e *a fortiori* ai sistemi di intelligenza artificiale

⁶⁵ I. Asimov, “The evitable conflict”, Id., in *I, robot*, Gnome Press, New York, 1950.

⁶⁶ S.M.C. Avila Negri, “Robot as Legal Person: Electronic Personhood in Robotics and Artificial Intelligence”, in *Frontiers in robotics and AI*, 8 (2021), 789327 (<https://doi.org/10.3389/frobt.2021.789327>).

⁶⁷ M. Farina, “Brevi riflessioni sullo status delle ‘persone elettroniche’”, in *L’Ircocervo*, 20 (2021), n. 2, pp. 106-126.

⁶⁸ A. Narayan, D. Perrott, “Can computers have legal rights?”, in M. Yazdani, A. Narayanan (eds.), *Artificial Intelligence. Human Effects*, Ellis Horwood Series, Halsted Pr., New York, (1984), pp. 52-61.

del nuovo millennio⁶⁹, giustificando così una nuova, seria ed imprevista generazione di diritti e doveri, riferibili a persone umane e soggettività artificiali, che dovranno abitare insieme nel nuovo “tecnomondo”.

⁶⁹ B.J. Koops, M. Hildebrandt, D.O. Jacquet-Chiffelle, “Bridging the Accountability Gap: Rights for New Entities in the Information Society?”, in *Minnesota Journal of Law, Science & Technology*, (2010), pp. 497-561.