amministrativ@mente
Rivista scientifica trimestrale di diritto amministrativo
www.amministrativamente.com

UNIVERSITÀ DEGLI STUDI DI ROMA "FORO ITALICO"

Rivista scientifica trimestrale di diritto amministrativo (Classe A)

Pubblicata in internet all'indirizzo www.amministrativamente.com

Rivista di Ateneo dell'Università degli Studi di Roma "Foro Italico"

Direzione scientifica
Gennaro Terracciano, Gabriella Mazzei, Julián Espartero Casado

Direttore Responsabile
Gaetano Caputi

Redazione
Giuseppe Egidio Iacovino, Carlo Rizzo

FASCICOLO N. 1/2024

Estratto

Iscritta nel registro della stampa del Tribunale di Roma al n. 16/2009

ISSN 2036-7821

amministrativ@mente
Rivista scientifica trimestrale di diritto amministrativo
www.amministrativamente.com

Comitato scientifico

Annamaria Angiuli, Antonio Barone, Vincenzo Caputi Jambrenghi, Francesco Cardarelli, Enrico Carloni, Maria Cristina Cavallaro, Guido Clemente di San Luca, Andry Matilla Correa, Chiara Cudia, Gianfranco D'Alessio, Mariaconcetta D'Arienzo, Ambrogio De Siano, Ruggiero Dipace, Luigi Ferrara, Pierpaolo Forte, Gianluca Gardini, Biagio Giliberti, Emanuele Isidori, Bruno Mercurio, Francesco Merloni, Giuseppe Palma, Alberto Palomar Olmeda, Attilio Parisi, Luca Raffaello Perfetti, Fabio Pigozzi, Alessandra Pioggia, Helene Puliat, Francesco Rota, Josè Manuel Ruano de la Fuente, Leonardo J. Sańchez-Mesa Martĩnez, Ramón Terol Gómez, Antonio Felice Uricchio.

Comitato editoriale

Jesús Avezuela Cárcel, Giuseppe Bettoni, Sveva Bocchini, Salvatore Bonfiglio, Vinicio Brigante, Sonia Caldarelli, Giovanni Cocozza, Andrea Marco Colarusso, Sergio Contessa, Giuseppe Doria, Manuel Delgado Iribarren, Fortunato Gambardella, Flavio Genghi, Jakub Handrlica, Margherita Interlandi, Laura Letizia, Giuseppina Lofaro, Federica Lombardi, Gaetano Natullo, Carmen Pérez González, Giovanni Pesce, Marcin Princ, Sara Pugliese, Antonio Saporito, Giuliano Taglianetti, Simona Terracciano, Stefania Terracciano, Salvatore Villani.

Coordinamento del Comitato editoriale

Valerio Sarcone.

ISSN 2036-7821

amministrativ@mente
Rivista scientifica trimestrale di diritto amministrativo
www.amministrativamente.com

Intelligenza Artificiale, tra profili di responsabilità e protezione dei dati personali: aspetti *de jure condito* e prospettive *de jure condendo*

di Filippo Lorè

(Professore a contratto presso l'Università degli Studi di Bari "A. Moro")

e

di Paolo Musacchio

(Avvocato e Data Protection Officer)

Sommario

1. Introduzione. – 2. Il quadro normativo rilevante in materia di IA. 2.1 Brevi cenni al c.d. caso GPT e alla posizione del Garante per la protezione dati personali. – 3. La responsabilità civile dell'IA. 3.1 Autonomia dei sistemi algoritmici e regolazione dei rapporti. 3.2 Le diverse teorie della dottrina sulla responsabilità dell'IA. 3.3 L'applicazione delle norme sulla responsabilità oggettiva. 3.4 I principi della Proposta di Direttiva UE sulla responsabilità extracontrattuale dell'IA. – 4. IA in medicina. – 5. Aspetti normativi in materia di tutela dei dati personali utilizzati nell'IA: l'*explainability* dell'algoritmo. – 6. Conclusioni

Abstract

In this contribution we intend to analyze, through a necessary regulatory examination, the most complex aspects that can arise in the use of Artificial Intelligence techniques. There is an ongoing challenge that permeates all sectors of daily life and, therefore, calls the public and private world to activate suitable guarantee mechanisms to safeguard the fundamental rights and freedoms of citizens, even more so when these systems are used in healthcare.

* *Il presente lavoro è stato sottoposto al preventivo referaggio secondo i parametri della double blinde peer review*

1. Introduzione

Sono necessarie, anzitutto, alcune premesse sotto il profilo definitorio, poiché Intelligenza artificiale (IA), *Machine Learning* (*ML*) e *Deep Learning* (*DL*) vengono a volte usati, impropriamente, come sinonimi. L'IA, campo dell'informatica dedicato alla creazione di sistemi che svolgono compiti che normalmente richiederebbero l'intelligenza umana, ricomprende il *ML*, ossia quegli approcci che permettono alle macchine di imparare dai dati, senza che queste siano programmate esplicitamente, dunque, con l'apprendimento, gli algoritmi di *ML* tenteranno di minimizzare gli errori e massimizzare la probabilità che le loro previsioni siano vere[1]. Il *DL* è un sottogruppo del *ML* e incorpora modelli e algoritmi computazionali che imitano l'architettura delle reti neurali biologiche nel cervello (reti neurali artificiali - *ANN*)[2]. Ciò detto, secondo un recente studio di Confartigianato, l'Intelligenza Artificiale (IA) metterebbe a rischio, nel nostro Paese, 8,4 milioni di lavoratori, ossia il 36,2%, per lo più nelle professioni maggiormente qualificate[3]. Tuttavia, siamo ormai di fronte a uno dei temi principali della riflessione filosofica e giuridica non tanto per i prevedibili problemi di occupazione derivanti dall'uso dell'IA, quanto per via dei potenziali impatti negativi sulla società contemporanea e, dunque, sulle persone, in special modo in presenza di decisioni automatizzate dei sistemi algoritmici o laddove i dati raccolti dall'IA siano relativi all'orientamento sessuale o alle convinzioni religiose, che possono alimentare *bias* negativi[4]. Per *bias* si intende ogni ragione che porta l'algoritmo a un risultato diverso da quello cui dovrebbe condurre il suo normale funzionamento[5]. I *bias* possono essere presenti nei *training data*, nella struttura dell'algoritmo, così come nell'utilizzo dell'algoritmo in un contesto diverso da quello per cui è stato concepito, nonché negli esiti dell'algoritmo laddove esso

---

* Benché il presente lavoro sia frutto della riflessione comune e della collaborazione dei due Autori, tuttavia, i paragrafi 3, 4 e 6 sono stati redatti da Filippo Lorè ed i paragrafi 1, 2 e 5 sono stati redatti da Paolo Musacchio.

[1] D. Jakhar, I. Kaur, *Artificial intelligence, machine learning and deep learning: definitions and differences*, in *Clin. Exp. Dermatol*, 2019.

[2] Y. LeCun, Y. Bengio, G. Hinton, *Deep learning*, in *Nature*, 512, 2015, pp. 436-444.

[3] Per Confartigianato l'AI mette a rischio 8,4 milioni di lavoratori in Italia, in *AI4Business*, 24 agosto 2023, URL https://www.ai4business.it/intelligenza-artificiale/per-confartigianato-lai-mette-a-rischio-84-milioni-di-lavoratori-in-italia/?utm_campaign=ai4business_20230825&utm_source=ai4business_20230825&utm_medium=email&sfdcid=0030O00002nYTGRQA4.

[4] S. Orlando, *Regole di immissione sul mercato e «pratiche di intelligenza artificiale» vietate nella proposta di Artificial Intelligence Act*, in *Persona e Mercato*, 3, 2022.

[5] B. Friedman, H. Nissenbaum, *Bias in computer systems*, in *ACM Transactions on Information Systems*, 3, 1996; J. Silberg, J. Manyika, *Notes from the AI frontier: Tackling bias in AI (and in humans)*, in *McKinsey Global Institute*, 2019.

riproduce *bias* esistenti nella società[6]. L'approccio più diffuso è, dunque, quello "antropocentrico", ossia focalizzato sui valori umani di cui dovrebbe farsi portatrice l'IA, dunque, prevedendo una soglia minima di controllo e sorveglianza umani (*human oversight*) sugli algoritmi, tuttavia, in tal senso viene sondata anche la possibilità di una prospettiva diversa – costituita da una «regolazione etica» dell'IA, preordinata ad attribuire una responsabilità etica delle azioni di tecnologie algoritmiche[7]. Considerata, infatti, la crescente diffusione dei sistemi automatizzati nei meccanismi decisionali[8], appare necessario prevenire eventuali rischi di "discriminazione sociale" derivanti dalle decisioni degli algoritmi, poiché «garantire che i diritti umani siano rafforzati e non minati dall'intelligenza artificiale è uno dei fattori chiave che definiranno il mondo in cui viviamo»[9]. Poiché gli algoritmi potrebbero orientarsi verso scopi non previsti in origine, sorge la necessità giuridica di allocare la responsabilità delle decisioni prese da tali sistemi[10], di talché, le Istituzioni di molti Paesi, tra cui anche quelle dell'Unione europea, stanno tentando, da qualche anno, di dare risposte, a garanzia del rispetto dei diritti fondamentali, da un lato, all'individuazione di un regime di responsabilità per le conseguenze dannose delle scelte dell'IA, e nel presente contributo si tenterà di passare in rassegna le diverse teorie in merito. D'altro canto, all'esigenza di individuare opportuni principi di *governance* dell'IA. Si pensi, oltre alla supervisione umana, a principi quali la sicurezza, la trasparenza, il benessere sociale e ambientale, la diversità, la non discriminazione e l'equità, la riservatezza dei dati. In particolare, il

[6] F. J. Zuiderveen Borgesius, *Strengthening legal protection against discrimination by algorithms and artificial intelligence*, in *The International Journal of Human Rights*, 10, 2020, p. 1572 ss.

[7] E. Chiti, B. Marchetti, *Divergenti? Le strategie di Unione europea e Stati Uniti in materia di intelligenza artificiale*, in *Rivista della Regolazione dei mercati*, 1, 2020.

[8] Nell'ampia letteratura in materia, si consigliano G. Phillips-Wren, L. Jain, *Artificial Intelligence for Decision Making*, in *Knowledge-Based Intelligent Information and Engineering Systems*, Berlino-Heidelberg, 2006, pp. 531–536; Y. R. Shrestha, S.M. Ben-Menahem, G. Von Krogh, *Organizational Decision-Making Structures in the Age of Artificial Intelligence*, in *California Management Review*, 61, 4, 2019, pp. 66–83; K. De Fine Licht, J. De Fine Licht, *Artificial Intelligence, Transparency, and Public Decision-Making: Why Explanations Are Key When Trying to Produce Perceived Legitimacy*, in *AI& SOCIETY*, 35, 4, 2020, pp. 917–926; A. Santosuosso, *Intelligenza Artificiale e Diritto: Perché Le Tecnologie Di IA Sono Una Grande Opportunità per Il Diritto*, Milano, 2020.

[9] Consiglio d'Europa, *Unboxing artificial intelligence: 10 steps to protect human rights*, 14 maggio 2019, https://www.coe.int/en/web/portal/news-2019/-/asset_publisher/gaVFYRTI7hqZ/content/unboxing-artificial-intelligence-10-steps-to-protect-human-rights?_101_INSTANCE_gaVFYRTI7hqZ_languageId=it_IT, in cui si afferma l'esigenza di prevedere una valutazione dell'impatto sui diritti umani (Human rights impact assessments, HRIA) da parte delle Autorità pubbliche che acquisiscono e sviluppano sistemi di IA, con particolare attenzione per i gruppi di persone che registrano un maggiore rischio di impatto dal ricorso all'IA.

[10] L. Savolainen, M. Ruckenstein, *Dimensions of autonomy in human–algorithm relations*, in *New Media & Society*, Giugno 2022, https://doi.org/10.1177/14614448221100802.

legislatore europeo ha affermato l'esigenza che i sistemi di IA collocati sul mercato prevedano, sin dalla fase di progettazione, l'adesione alla normativa in materia di tutela dei diritti fondamentali e ai valori dell'Unione, nonché «migliorare la governance e l'applicazione effettiva della normativa esistente in materia di diritti fondamentali e requisiti di sicurezza applicabili ai sistemi di IA»[11]. Tali considerazioni valgono ancor di più per l'IA generativa e per i sistemi algoritmici "predittivi". Questi ultimi sono sussumibili in diverse categorie, poiché si va dalla polizia predittiva, utilizzata per selezionare le zone delle città in cui la polizia dovrebbe essere presente a tutela dell'ordine pubblico, alla giustizia predittiva[12], alle predizioni del traffico, al *welfare allocation*, in grado di decidere se un cittadino è idoneo a beneficiare di un servizio pubblico, al *social scoring*, alla medicina predittiva, con evidenti problemi, in tutti i casi citati, di natura etica. In particolare, ci si soffermerà su alcuni risvolti dell'uso degli algoritmi nel campo della medicina.

2. Il quadro normativo rilevante in materia di IA

Nell'ambito della *governance* dell'Intelligenza Artificiale (IA)[13], in ambito europeo, il punto di partenza è rintracciabile in due atti della Commissione Europea, ossia la Comunicazione del 25 aprile 2018, dal titolo "*L'intelligenza artificiale per l'Europa*", e il rapporto "*Ethics guidelines for trustworthy AI*", a firma dello "*Independent high-level expert group on artificial intelligence*", i cui risultati sono stati ripresi il 19 febbraio 2020 nel Libro Bianco sull'intelligenza artificiale [14]. Le Istituzioni europee, mirando a conciliare lo sviluppo sicuro e affidabile dell'IA con il «pieno rispetto dei valori e dei diritti dei cittadini dell'UE», e a creare un «ecosistema di fiducia», auspicano che l'IA non solo rispetti la legge, ma osservi anche i principi etici, garantendo «che le loro attuazioni pratiche non comportino danni indesiderati», puntando a rafforzare «le

[11] D. Chiappini, *Intelligenza Artificiale e responsabilità civile: nuovi orizzontidi regolamentazione alla luce dell'Artificial Intelligence Actdell'Unione europea*, in *Rivista Italiana di Informatica e Diritto*, 2, 2022.

[12] In tal senso, cfr., *ex multis*, R. Bichi, *Intelligenza Artificiale tra "calcolabilità" del diritto e tutela dei diritti*, in *Giur. it.*, 2019, p. 1172; E. Battelli, *Giustizia predittiva, decisione robotica e ruolo del giudice*, in *Giust. civ.*, 2020, p. 281 ss.; A. Carratta, *Decisione robotica e valori del processo*, in *Riv. dir. proc.*, 2020, p. 491 ss.

[13] L'art. 3, AI Act, definisce, quale "sistema di intelligenza artificiale", un sistema progettato per funzionare con elementi di autonomia e che, sulla base di dati e input forniti da macchine e/o dall'uomo, deduce come raggiungere una determinata serie di obiettivi avvalendosi di approcci di apprendimento automatico e/o basati sulla logica e sulla conoscenza, e produce output generati dal sistema quali contenuti (sistemi di IA generativi), previsioni, raccomandazioni o decisioni, che influenzano gli ambienti con cui il sistema di IA interagisce.

[14] Commissione UE, "LIBRO BIANCO sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia", 2020, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_it.pdf

capacità dei cittadini, non a sostituirsi a loro, e consentire l'accesso anche alle persone con disabilità»[15]. Il Parlamento Europeo ha emanato il 20 ottobre 2020 tre risoluzioni, due proposte di regolamento su etica, responsabilità e proprietà intellettuale dei sistemi di IA e una quarta risoluzione sui sistemi autonomi di armi letali. Non va dimenticata, inoltre, la recente adozione del sistema di norme costituito dal *Digital Services Act* (*DSA*, legge sui servizi digitali) e dal *Digital Markets Act* (*DMA*, legge sul mercato digitale) che ha come obiettivi principali quello di creare uno spazio digitale più sicuro in cui siano tutelati i diritti fondamentali di tutti gli utenti dei servizi digitali e quello di creare condizioni di parità per promuovere l'innovazione, la crescita e la competitività, sia nel mercato unico europeo che a livello globale. Il DSA, in particolare, va a sostituire e a novare il precedente regime di responsabilità dei fornitori di servizi della società dell'informazione, costituito dalla Direttiva E-commerce (Dir. CE 2000/31), mirando a creare un'ambiente digitale più sicuro, in cui i diritti fondamentali degli utenti dei servizi digitali siano protetti, in particolare per fini di stabilità democratica, atteso il crescente potere di piattaforme ed esperti (*gatekeepers*), connesso al fenomeno di disinformazione e polarizzazione, mediante usi impropri dei servizi *online* da parte di sistemi algoritmici manipolativi[16]. Infine, nel 2021, la Commissione Europea ha approvato la Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale, *AI Act*) e modifica alcuni atti legislativi dell'Unione, 21 aprile 2021 (2021/0106 (COD)), di cui è prevista l'adozione definitiva, all'esito dei "triloghi" tra le istituzioni europee, entro la fine del 2023. La Proposta di Regolamento sul'IA, definita da Pizzetti come «snodo nel passaggio della UE dalla società predigitale a quella digitale»[17], stabilisce regole armonizzate direttamente applicabili su tutto il territorio europeo in materia di sviluppo, commercializzazione e utilizzo dell'IA, garantendo la libera circolazione di tali sistemi. L'esigenza primaria del legislatore europeo è quella di garantire che i sistemi di IA collocati sul mercato prevedano, sin dalla fase di progettazione, l'adesione alla normativa in materia di tutela dei diritti fondamentali e ai valori

[15] D. Iacovelli, M. Fontana, *Nuove sfide della tecnologia e gestione dei rischi nella proposta di regolamento europeo sull'intelligenza artificiale: set di training, algoritmi e qualificazione dei dati. Profili critici*, in *Il diritto dell'economia*, 109, 3, 2022, p. 106 ss.

[16] E. Garzonio, *L'algoritmo trasparente: obiettivi ed implicazioni della riforma dello Spazio digitale europeo*, in *Rivista italiana di informatica e diritto (RIID)*, 3, 2, 2021.

[17] F. Pizzetti, *La proposta di Regolamento sull'IA della Commissione Europea presentata il 21.4.2021 (COM (2021) 206 final) tra Mercato Unico e competizione digitale globale*, in *Diritto di internet*, 4, 2021.

dell'Unione, nonché «migliorare la governance e l'applicazione effettiva della normativa esistente in materia di diritti fondamentali e requisiti di sicurezza applicabili ai sistemi di IA»[18]. Poiché l'UE intende regolamentare l'IA in maniera proattiva e in base ai rischi potenziali (bassi, medi o elevati) per i diritti fondamentali (approccio *risk based*), le conseguenze per l'industria tecnologica, a livello globale, non sono di poco conto. Quanto più sarà alto il rischio, maggiori saranno gli adempimenti e le responsabilità degli autori delle applicazioni di IA, che potranno avvalersi, nel processo di *compliance*, dell'adozione di standard di riferimento predisposti dagli enti di normazione quali ISO e CEN, nonché dell'inserimento in un apposito registro pubblico. Un obiettivo delle istituzioni europee è la flessibilità della disciplina, attesa la continua e rapida evoluzione della società digitale e della tecnologia, sulla scia di quanto operato in materia di protezione dei dati personali con il Reg. UE 2016/679, improntato a un sistema di principi e meccanismi dinamici, destinati a mettersi in relazione a una realtà mutevole[19], e che ha introdotto, da un lato, l'obbligatorietà della valutazione di impatto per quei trattamenti che comportino rischi per i diritti e le libertà delle persone, in particolare quando siano utilizzate nuove tecnologie, d'altro canto, obblighi di consultazione dell'Autorità garante, qualora sussistano rischi residui elevati, nell'ottica dell'accountability del titolare del trattamento[20]. La Commissione nel prevedere, nell'art. 7, par. 2, dell'*IA Act*, un sistema di valutazione del rischio di danno per la salute e la sicurezza o di pregiudizio per i diritti fondamentali da parte di un sistema di IA, ha elencato una serie di criteri per stabilire la gravità del rischio dei sistemi di IA[21]: la finalità prevista del sistema di IA; la misura in cui un sistema di IA è stato usato o è probabile che sarà usato; la misura in cui l'uso di un sistema di IA ha già causato un danno alla salute e alla sicurezza, un impatto negativo sui diritti fondamentali o ha suscitato gravi preoccupazioni in relazione al verificarsi di tale danno o impatto negativo; la portata potenziale di tale danno o di tale impatto negativo; la misura in cui le persone che potrebbero subire il danno o l'impatto negativo dipendono dal risultato prodotto da un sistema di IA; la misura in cui le persone che potrebbero subire il danno o l'impatto negativo si trovano in una

[18] D. Chiappini, *Intelligenza Artificiale…, op. cit.*
[19] F. Pizzetti, *La proposta…, op. cit.*
[20] D. Iacovelli, M. Fontana, *Nuove sfide…, op. cit.*
[21] Commissione UE, Proposta di Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni atti legislativi dell'Unione, 21 aprile 2021 (2021/0106 (COD))

posizione vulnerabile rispetto all'utente di un sistema di IA; la misura in cui il risultato prodotto con un sistema di IA è facilmente reversibile; la misura in cui la legislazione vigente dell'UE prevede misure di ricorso efficaci in relazione ai rischi presentati da un sistema di IA, ad esclusione delle richieste di risarcimento del danno, e misure efficaci per prevenire o ridurre sostanzialmente tali rischi[22]. Il terzo allegato dell'*AI Act*, dunque, riporta un elenco di sistemi di IA ad alto rischio, in considerazione di eventi che si sono già concretizzati o potrebbero verificarsi nel prossimo futuro. Tale elenco è sempre ampliabile dalla Commissione, e l'art. 84 prevede un suo riesame annuale, mentre, ogni quattro anni, la Commissione è tenuta a trasmettere al Parlamento e al Consiglio una relazione di valutazione e riesame dell'*AI Act*, nonché una valutazione di impatto e di efficacia dei codici di condotta[23]. Come osservato in dottrina, in una sorta di "recupero a valle" della dimensione del principio di proporzionalità, dalle più recenti versioni dell'*IA Act* emerge una certa flessibilità, laddove si prevede che, anche per i sistemi ad alto rischio, debba prendersi in considerazione la significatività dell'impatto rischioso dell'output del sistema di IA, mettendolo in relazione alla decisione o all'azione intrapresa da una persona fisica[24]. Sebbene la maggior parte delle applicazioni di IA presenti un rischio basso, vi sono dei sistemi algoritmici che devono essere trasparenti, affinché appaia evidente l'assenza di una supervisione umana (si pensi alle *chatbot* di supporto dell'utente in un sito *web*). Il rischio è alto nelle applicazioni dalle cui decisioni possono derivare conseguenze sulla vita delle persone, quali, ad esempio, l'ottenimento di un prestito, l'assunzione in una posizione lavorativa, la circolazione di un'auto a guida autonoma e l'uso di un dispositivo medico; per questo motivo il legislatore introduce obblighi riguardanti la condivisione con gli utenti e con le Autorità competenti, di documentazione esaustiva e dettagliata sul funzionamento dei sistemi, un adeguato livello di supervisione umana nella progettazione e nell'uso dei sistemi, la qualità dei *dataset* e il rispetto dei più alti standard di accuratezza e di sicurezza informatica[25]. Su diversi aspetti della disciplina, tuttavia, non si è raggiunto un accordo, e sono stati numerosi i rallentamenti del processo legislativo di adozione del Regolamento. Da ultimo, si registra il tentativo, nel giugno 2023, di un gruppo di

[22] D. Chiappini, *Intelligenza Artificiale…, op. cit.*

[23] D. Iacovelli, M. Fontana, *Nuove sfide della tecnologia e gestione dei rischi nella proposta di regolamento europeo sull'intelligenza artificiale: set di training, algoritmi e qualificazione dei dati. Profili critici*, in *Il diritto dell'economia*, 109, 3, 2022, pp. 106-138.

[24] G. Lo Sapio, *Intelligenza artificiale: rischi, modelli regolatori, metafore*, in *federalismi.it*, 27, 2022.

[25] S. Orlando, *Regole di immissione…, op. cit.*

parlamentari europei, di escludere il riconoscimento biometrico dalle pratiche vietate dal Regolamento. Va altresì rilevato che le norme del Regolamento si applicano ai fornitori di sistemi di IA, indipendentemente dal fatto che siano stabiliti nell'Unione o in un paese terzo, agli utenti dei sistemi di IA stabiliti nell'Unione e ai fornitori e utenti dei sistemi di IA, stabiliti in un paese terzo al di fuori dell'Unione, nella misura in cui i sistemi di IA interessano le persone situate nell'Unione. L'*AI Act* stabilisce, all'articolo 1: (i) regole armonizzate per l'immissione sul mercato, la messa in servizio e l'uso dei sistemi di IA; (ii) il divieto di determinate pratiche di intelligenza artificiale; (iii) requisiti specifici per i sistemi di IA ad alto rischio e obblighi per gli operatori di tali sistemi; (iv) regole di trasparenza armonizzate per i sistemi di IA destinati a interagire con le persone fisiche, i sistemi di riconoscimento delle emozioni, i sistemi di categorizzazione biometrica e i sistemi di IA utilizzati per generare e manipolare immagini, contenuti audio o video; (v) regole in materia di monitoraggio e vigilanza del mercato. In dottrina è stato affermato che il Regolamento sull'IA costruisce «un "processo" che consente di gestire i rischi connessi all'uso dell'IA al fine di prevenire pregiudizi, non regolando la materia in maniera verticale (per *silos* normativi), ma in modo orizzontale»[26]. Il Parlamento europeo ha esteso il divieto sui *software* di identificazione biometrica, introdotto un livello aggiuntivo, ad alto rischio, per i sistemi che possono provocare danni alla salute, alla sicurezza o ai diritti fondamentali e per i sistemi di raccomandazione delle piattaforme online di grandi dimensioni. Con riferimento al controllo predittivo, il divieto è stato esteso dai reati penali a quelli amministrativi. Il Regolamento intende sbarrare la strada a usi dell'IA per finalità considerate contrarie ai valori dell'UE, come, ad esempio, il *social scoring*, e le applicazioni che sfruttano le vulnerabilità di gruppi di persone a causa della loro situazione sociale o economica[27]. Regole severe sono presenti, nel nuovo testo, per i cd. *foundation model*, quali *ChatGPT*, per gli strumenti di IA per il monitoraggio generale delle comunicazioni, nonché per i *software* di riconoscimento delle emozioni alimentati dall'AI, vietato nei settori dell'applicazione della legge, della gestione delle frontiere, del lavoro e dell'istruzione. In particolare, sistemi di *machine learning*, come *ChatGPT*, non controllando, durante il processo di "apprendimento", le proprie fonti, sono inclini a favorire la diffusione di opinioni infondate, pregiudizi e *fake news*. L'attuale bozza di

[26] M. Rabitti, *Il danno prodotto dall'algoritmo intelligente: quale modello di responsabilità civile?*, in *Trimestrale dell'Associazione Donne magistrato Italiane, giudicedonna.it*, 3-4, 2021.

[27] D. Chiappini, *Intelligenza Artificiale…, op. cit.*

Regolamento consente agli Stati membri di individuare l'autorità nazionale sull'intelligenza artificiale. Per quanto concerne l'impianto sanzionatorio, per i casi di violazioni dell'*AI Act*, le autorità potranno comminare sanzioni fino a 30 milioni di euro o al 6% del fatturato mondiale annuo (ma, per le PMI, al 3%). Alla luce delle criticità dell'IA sul piano etico, che verranno trattate nel prosieguo di questo contributo, la disciplina che il legislatore europeo intende introdurre, come osservato in dottrina, non rappresenta un eccesso, bensì il nucleo minimo delle cautele necessarie per l'utilizzo dei sistemi algoritmici[28]. Vengono proibiti, poiché considerati come "inaccettabili" per la Commissione Europea, i sistemi di IA che usano tecniche subliminali per arrecare danno fisico o psicologico alle persone, quale, ad esempio, un assistente vocale di un giocattolo che manipola la mente di un bambino e lo induce a un comportamento pericoloso, i sistemi di assegnazione di un punteggio "sociale" (*social scoring*), che classificano le persone in base al comportamento sociale, la videosorveglianza di massa, ovvero l'identificazione biometrica remota nei luoghi pubblici[29]. A garanzia del rispetto dei diritti fondamentali, si pongono alcuni principi di *governance* come la supervisione umana (*human oversight*), la sicurezza, la trasparenza, il benessere sociale e ambientale, la diversità, la non discriminazione e l'equità, la riservatezza dei dati, in special modo laddove i dati raccolti siano relativi all'orientamento sessuale o alle convinzioni religiose, che possono alimentare *bias* negativi[30]. I produttori, dunque, dovranno dimostrare che i servizi basati sull'IA non ledono i diritti delle persone, *in primis* con l'assenza di possibili discriminazioni degli algoritmi, e, poi, con la garanzia del diritto all'accesso all'informazione. La Commissione UE, con i Consideranda 53 e 58 alla Proposta di Regolamento sull'IA, ha inteso stabilire che, da un lato, il fornitore (che sia persona fisica o giuridica) del prodotto intelligente si debba assumere la responsabilità dell'immissione sul mercato o della messa in servizio di un sistema di IA ad alto rischio, a prescindere dal fatto che egli sia la persona che ha progettato o sviluppato il sistema; d'altra parte, invece, sancisce che è opportuno stabilire responsabilità specifiche per gli utenti e, in particolare, «che gli utenti usino i sistemi di IA ad alto rischio conformemente alle istruzioni per l'uso e che siano previsti alcuni altri obblighi in materia di

[28] D.-U. Galetta, *Human-stupidity-in-the-loop? Riflessioni (di un giurista) sulle potenzialità e i rischi dell'Intelligenza Artificiale*, in *federalismi.it*, 5, 2023.

[29] Roundtable of G7 Data Protection and Privacy Authorities, "Promoting Data Free Flow with Trust and knowledge sharing about the prospects for International Data Spaces", 8 settembre 2022, https://www.bfdi.bund.de/SharedDocs/Downloads/EN/G7/Communique-2022.pdf?__blob=publicationFile&v=1

[30] S. Orlando, *Regole di immissione…, op. cit.*

monitoraggio del funzionamento dei sistemi di IA e conservazione delle registrazioni, a seconda dei casi»[31]. Nell'ottica dell'accennata *accountability*, per i sistemi ad alto rischio, i risultati dovranno essere verificati e tracciati nel loro intero ciclo di vita e la loro documentazione tecnica dovrà essere redatta prima dell'immissione sul mercato, aggiornata di continuo, con la relativa conservazione obbligatoria dei *file* di *log* degli eventi (art. 12). Di talché, i fornitori saranno tenuti a dimostrare all'autorità la "conformità" dei sistemi "ad alto rischio", che devono essere progettati e sviluppati, in modo tale da garantire la trasparenza, implementati con strumenti che ne consentano la supervisione umana, finalizzata a prevenire e ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali (art. 14)[32]. Inoltre, essi devono mantenere un adeguato livello di accuratezza, robustezza e cibersicurezza che duri per l'intero ciclo di vita (art. 15). Con riferimento ai sistemi ad alto rischio, si segnala il recente report *Cybersecurity of Artificial Intelligence in the IA Act*, uno studio di alto livello che si concentra sul requisito di sicurezza informatica per i sistemi IA ad alto rischio, come stabilito nell'articolo 15 della proposta della Commissione europea per l'*AI Act*, attraverso quattro principi guida chiave per ottenere la conformità all'*AI Act*[33]. Per contro, sono state segnalate delle criticità in merito, da un lato, alla rigidità delle definizioni del Regolamento, e, dall'altro, alla carenza di dinamicità rispetto a eventuali futuri sviluppi della tecnologia[34]. Tali inadeguatezze sono dovute alla scelta di procedere con un approccio *top-down*, ovvero è il legislatore a individuare, *ex ante*, il diverso livello di rischio, dal quale far discendere le regole applicabili, con il pericolo di elaborare categorie normative astratte e poco aderenti alla realtà in divenire. Sarebbe stato più opportuno optare

[31] Commissione UE, Proposta di Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni atti legislativi dell'Unione, 21 aprile 2021 (2021/0106 (COD))

[32] G. D'Alfonso, *Danni algoritmici e sviluppi normativi europei tra "liability" e "permittance" rules*, in *EJPLT*, 18-66, Suor Orsola Benincasa Università Editrice, 2022.

[33] H. Junklewitz, R. Hamon, A. André, T. Evas, J. Soler Garrido, J. Sanchez Martin (Commissione UE), *Cybersecurity of Artificial Intelligence in the IA Act, Guiding principles to address the cybersecurity requirement for high-risk AI systems*, settembre 2023, in cui i quattro principi guida sono: a) il focus dell'AI Act è sull'intero sistema di IA e non direttamente ai suoi componenti interni; b) la conformità all'*AI Act* richiede necessariamente una valutazione dei rischi di sicurezza, prendendo in considerazione l'architettura interna del sistema di AI e il contesto di applicazione previsto al fine di identificare i rischi specifici, tradurre il requisito generale del regolamento in requisiti specifici per i componenti del sistema e implementare le necessarie misure di mitigazione; c) la sicurezza dei sistemi di IA richiede un approccio integrato e continuo (olistico) che utilizzi pratiche comprovate e controlli specifici per l'IA; d) non tutte le tecnologie di IA potrebbero essere pronte per l'uso in in scenari ad alto rischio se le loro limitazioni in termini di cybersecurity non siano affrontate adeguatamente.

[34] G. D'Alfonso, *Ibidem*.

per un approccio *bottom-up*, simile a quello del *GDPR*, con un apprezzamento del rischio conferito a una valutazione decentralizzata[35]. D'altro canto, vi è in dottrina chi muove, al legislatore europeo, l'accusa di eccessiva discrezionalità lasciata agli Stati membri e una certa indeterminatezza, con norme che paventano abusi e pericoli, sanzionano comportamenti, anche se in alcuni casi "solo in linea di principio" (*id est*, *social scoring*) o, in altri casi (identificazione biometrica real time in spazi aperti al pubblico) salvo che non risultino strettamente necessari per la ricerca mirata di potenziali condotte criminose, ovvero per la prevenzione di un pericolo specifico; il rischio risiede nella discrezionalità del potere esercitato tramite l'IA, che potrebbe degenerare in arbitrio[36]. Da tali riflessioni, appare evidente la centralità, nella regolazione delle decisioni algoritmiche, del tema della responsabilità dei sistemi di IA. Il principio di "*liability* e *legal responsibility*" si riferisce, infatti, all'esigenza di garantire che le persone o le entità responsabili del danno causato da un sistema di IA possano essere ritenute responsabili.

2.1 Brevi cenni al c.d. caso GPT e alla posizione del Garante per la protezione dati personali

Alla luce di quanto suesposto, preme, seppur brevemente, riportare un caso di studio. Quanto all'utilizzo dei sistemi di IA, non si può prescindere dal considerare gli effetti, lato *privacy*, verificatisi per il caso c.d. *ChatGPT* (*Generative Pre-trained Transformer*), ossia uno dei più diffusi tra i *software* di intelligenza artificiale relazionale, capace di simulare e, ulteriormente, di elaborare le conversazioni umane, mediante un sistema di dialogo basato su un *transformer* generativo pre-addestrato[37]. Tale strumento di elaborazione risponde agli *input* in un linguaggio naturale, ma riesce, altresì, a completare o suggerire testi sulla base di input parziali, dunque è utile per generare descrizioni, riassumere testi, operare traduzioni multilingue, nonché analizzare il *sentiment* del testo, consentendo di determinare se lo stesso testo

[35] G. Finocchiaro, *La regolazione dell'intelligenza artificiale*, in *Rivista trimestrale di diritto pubblico*, 4, 2022, pp. 1085-1099.

[36] T. E. Frosini, *L'orizzonte giuridico dell'intelligenza artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2022; C. Casonato, B. Marchetti, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di Intelligenza Artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2021.

[37] I. Coppola, *Intelligenza artificiale generativa: GPT o chatGPT verso ipotesi di ragionamento automatici e linguaggio artificiale. Tra l'argomentazione giuridica di Bobbio ed il principio di precauzione*, in *Diritto di internet*, 30 marzo 2023, in cui si afferma che il prototipo è stato sviluppato da OpenAI, ente no profit fondato da Elon Musk nel 2015, e che, per analizzare il dataset di Chat GPT sono stati utilizzati 175 miliardi di parametri (un parametro è la connessione tra un neurone e l'altro della rete neurale alla base del modello stesso).

è positivo, negativo o neutro[38]. Numerosi sono i punti di criticità, "lato privacy", di siffatto sistema, non ultimi quelli relativi alla trasparenza[39] e alla sicurezza del trattamento. Invero, nel marzo scorso, l'applicativo aveva registrato una perdita di dati (*c.d. data breach*), relativa alle conversazioni degli utenti e alle informazioni collegate al pagamento degli abbonati al servizio[40]. In merito, L'Autorità, contestualmente all'apertura di un'istruttoria, ha ordinato, con effetto immediato, la limitazione provvisoria del trattamento delle informazioni personali degli utenti italiani nei confronti di *OpenAI L.L.C.*, società, titolare del trattamento e con rappresentante all'interno dello Spazio Economico Europeo, protagonista della fase di sviluppo e di gestione della piattaforma. Nel Provvedimento in esame, il Garante per la protezione dei dati personali ha rimarcato, in primo luogo, l'inosservanza dell'art. 5, Reg. UE 2016/679, e, più precisamente, del principio di trasparenza, dove si è riscontrata l'assenza di una nota informativa, rivolta agli utenti e a tutti gli interessati, a firma della società *OpenAI L.L.C.* Ancor più, si è evidenziata l'inidonea individuazione di una base giuridica che giustifichi le rilevantissime operazioni di trattamento dati personali (dalla fase di raccolta a quella di conservazione), finalizzate ad "addestrare" gli algoritmi e funzionali alla corretta esecuzione dell'applicativo. All'esito delle verifiche condotte, si è evinto, altresì, che le informazioni fornite da *ChatGPT* non coincidono con il "dato reale", pertanto si configura un trattamento di dati non rispondente al principio di esattezza. n ultimo, dalla lettura delle condizioni pubblicate da *OpenAI L.L.C.* si ricavava che il servizio, pur essendo orientato agli interessati con età maggiore rispetto a tredici anni, non disponeva di alcuna misura tecnica deputata all'accertamento dell'età degli utenti e, dunque, tale carenza poneva in serio pericolo i minori, ai quali era consentito il libero accesso a tale strumento. Manifesta risultava, invero, la violazione del Reg. UE 2016/679 e del D.lgs n. 196/2003 e s.m.i., nella parte in cui si riconosce agli ultraquattordicenni la possibilità di consentire al trattamento per i servizi della società dell'informazione. Senonché, con Provvedimento dedicato[41], il Garante per la protezione dei dati personali, tenuto conto della collaborazione offerta dalla società *OpenAI L.L.C.* a collaborare con gli Uffici, ha inteso "*rivalutare la sussistenza dei*

[38] R. Dale, *GPT-3: What's it good for? in Natural Language Engineering*, in *Cambridge University*, vol. 27, gennaio 2021, pp. 113-118.

[39] Si rinvia, per tali aspetti, all'esaustiva trattazione di E. Prosperetti, *ChatGPT ed altri tool di AI. Condizioni di utilizzo, questioni di proprietà intellettuale e privacy*, in *Diritto di internet*, 2, 2023.

[40] Garante per la protezione dei dati personali, Provvedimento del 30 marzo 2023, doc. web n. 9870832.

[41] Garante per la protezione dei dati personali, Provvedimento dell'11 aprile 2023, doc. web n. 9874702.

presupposti del provvedimento di limitazione provvisoria", previa immediata attuazione di una serie di prescrizioni. In prima battuta, anche in adesione alle linee guida, *ratione materiae*, diramate dall'ormai ex *Gruppo di Lavoro Art. 29*[42], L'Autorità ha intimato di rendere nota, sul sito web, l'informativa *privacy*, rispondente all'art. 12, Reg. UE 2016/679, preordinata ad illustrare agli interessati (anche diversi dagli utenti del servizio *ChatGPT*), i cui dati sono stati raccolti e trattati ai fini dell'addestramento degli algoritmi, le modalità del trattamento, la logica alla base del trattamento necessario al funzionamento del servizio, vieppiù i diritti loro spettanti in qualità di interessati e ogni altra informazione prevista dalla disciplina rilevante in materia. Analogamente, la stessa Autorità ha segnalato, alla Società summenzionata, la necessità di implementare un *link* all'informativa privacy, indirizzato agli utenti dei servizi che si connettono dal territorio nazionale, in maniera tale da consentire la lettura del documento a favore di tutti coloro i quali intendono procedere alla registrazione sulla piattaforma o che desiderano riattivare l'iscrizione al servizio. Ulteriormente, sempre a tutela degli interessati, il Garante ha imposto l'adozione di uno strumento, mediante il quale riconoscere, agli utenti che si collegano dall'Italia e anche qualora la base giuridica sia rappresentata dal legittimo interesse, l'esercizio del diritto di opposizione, rispetto ai trattamenti dei propri dati personali, ottenuti da terzi, svolti dalla società ai fini dell'addestramento degli algoritmi e dell'erogazione del servizio. Ancora, l'Autorità prescrive l'adozione di misure organizzative tese ad assicurare agli utenti la richiesta e l'ottenimento della "*correzione di eventuali dati personali che li riguardano trattati in maniera inesatta nella generazione dei contenuti o, qualora ciò risulti impossibile allo stato della tecnica, la cancellazione dei propri dati personali*". Quanto alla definizione della base giuridica del trattamento dei dati personali degli utenti ai fini dell'addestramento degli algoritmi, il Garante prescrive di rimuovere ogni riferimento all'art. 6, par. 1, lett. b), Reg. UE 2016/679, individuando, piuttosto, in forza del principio di *accountability*, quale condizione di liceità, il consenso o il legittimo interesse. Per quanto attiene il controllo sull'età dei minori, l'Autorità ha disposto, oltre all'attivazione di un meccanismo di richiesta dell'età, per le finalità correlate alla registrazione del servizio, la realizzazione di un sistema di *age verification*, utile a inibire l'accesso agli utenti infratredicenni e ai minorenni, in assenza del consenso rilasciato dai genitori. Preme rilevare che l'azione del Garante, mirata non a fermare lo sviluppo dei sistemi di intelligenza artificiale,

[42] Gruppo di Lavoro Articolo 29 per la protezione dei dati, Linee guida sulla trasparenza ai sensi del regolamento 2016/679.

bensì a riflettere sulle logiche che sorreggono l'algoritmo e sulle relative implicazioni etiche, ha aperto la strada ad una task force su *ChatGPT* lanciata dal Comitato europeo per la protezione dati.

3. La responsabilità civile dell'IA

3.1 Autonomia dei sistemi algoritmici e regolazione dei rapporti

Un certo grado di autonomia può caratterizzare le tecniche basate sull'IA, intendendosi, come autonomo, quel sistema che opera interpretando determinati dati forniti, nonché utilizzando una serie di istruzioni predeterminate, senza essere limitato a tali istruzioni, nonostante il comportamento del sistema sia legato e volto al conseguimento dell'obiettivo impartito dallo sviluppatore in sede di progettazione[43]. L'aspetto più problematico è costituito dal fatto che potrebbero non essere ricondotte agli umani le ragioni delle decisioni prese da tali sistemi, nonché dalla possibilità che gli algoritmi si orientino verso scopi non previsti in origine[44], come d'altronde ha evidenziato il Parlamento UE nella Carta della Robotica del 2017, laddove ha affermato che l'autonomia di un *robot* può essere definita come la capacità di prendere decisioni e metterle in atto nel mondo esterno, indipendentemente da un controllo o un'influenza esterna[45], lì dove è stato raccomandato dalla Commissione di elaborare una definizione di «sistemi ciberfisici, di *robot* autonomi intelligenti e delle loro sottocategorie»[46], basata su diversi requisiti, tra cui quello dell'autonomia, oltre a quelli di autoapprendimento, esistenza di un supporto fisico, capacità di adattamento del proprio comportamento all'ambiente e assenza di vita, in termini biologici. L'autonomia degli algoritmi, che decidono sulla base di informazioni raccolte in precedenza che possono seguire regole spesso poco comprensibili, è, al tempo stesso, un punto di forza, e una debolezza dell'IA, considerato che tali regole, a volte inglobano fondamenti matematici o statistici, altre volte, invece, sono quelle stabilite da chi li ha programmati, dunque, sono frutto di

[43] Art. 4 "Proposta di Regolamento del Parlamento europeo e del Consiglio sui principi etici per lo sviluppo, la diffusione e l'utilizzo dell'intelligenza artificiale, della robotica e delle tecnologie correlate", 6 ottobre 2021, C 404/89.

[44] L. Savolainen, M. Ruckenstein, *Dimensions of autonomy…, op. cit.*

[45] Parlamento UE, Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).

[46] A. Santosuosso, C. Boscarato, F. Caroleo, *Robot e diritto: una prima ricognizione*, in *Nuova giur. civ. comm.*, 2012, p. 498, in cui è operata una tripartizione dei *robot* sulla base della loro autonomia (*robot* tele-operati, *robot* autonomi e *robot* cognitivi).

arbitrio[47]. D'altronde, la tecnologia è sempre meno un «mezzo» per le azioni umane e sempre più è essa stessa a prendere decisioni rilevanti per la persona umana e la sua libertà, poiché alla macchina non si chiede di realizzare ciò che un soggetto ha deciso, ma le si chiede di decidere, autonomamente, e, di conseguenza, «la causa precede irrevocabilmente e non può divenire posteriore al suo effetto»[48]. Invero, se si considera che l'IA consente di sostituire l'uomo in numerosi ambiti – quali la medicina, i trasporti, la finanza, l'uso delle armi, la *cybersecurity* – svolgendo in autonomia compiti che comporterebbero l'assunzione di responsabilità morali da parte di un esecutore umano, appaiono emergere anche delle questioni etiche, che verranno approfondite nel prosieguo della presente trattazione. Si pensi alla scelta che deve compiere il sistema che controlla una *self driving car* che si trova improvvisamente davanti tre persone che hanno iniziato ad attraversare la strada, con due sole possibilità, ossia investire i pedoni o sterzare bruscamente, portando il veicolo fuori strada e mettendo a repentaglio il passeggero, scelta, questa, che ha implicazioni significative sul piano delle responsabilità, dei doveri morali e dei diritti fondamentali delle persone. Non a caso, vi è chi ha avanzato l'idea di adottare una politica di segregazione per i veicoli autonomi su strade a essi riservate imponendo distanze interveicolari e altre misure di sicurezza tali da escludere l'insorgere di conflitti morali[49]. Da tali riflessioni derivano, dunque, conflitti etici che si ripercuotono anche sulle scelte di progettazione dei veicoli autonomi e che possono, a loro volta, generare conflitti tra gruppi sociali caratterizzati da diverse opzioni morali[50]. In tale scenario, da molti anni, in dottrina, ci si chiede se siano sufficienti le norme tradizionali per attivare la responsabilità per i danni causati dai sistemi di IA (in special modo i *robot*) «in quanto non consentirebbero di determinare qual è il soggetto cui incombe la responsabilità del risarcimento né di esigere da tale soggetto la riparazione dei danni causati»[51]. Le considerazioni richiamate sono ancor più pregnanti nel settore sanitario, non soltanto per quanto riguarda il rapporto tra

[47] D. Bennato, *Il computer come macroscopio. Big data e approccio computazionale per comprendere i cambiamenti sociali e culturali*, Franco Angeli, Milano, 2015.

[48] B. Romano, *Dalla metropoli verso Internet: persona libertà diritto. Appendice: Riflessioni su P. Grossi*, in *Ritorno al diritto*, edizione curata da V. De Felice e M. Innocenzi, Torino, 2017, p. 27 ss., in cui fa riferimento a M. Heidegger, *Dell'essenza della libertà umana: introduzione alla filosofia*, a cura di M. Pietropaoli, Milano, 2016, p. 397 ss.

[49] F. Santoni de Sio, *Ethics and self-driving cars. A White paper on responsible innovation in automated driving systems*, Rapporto interno della Delft University of Technology, 2016.

[50] F. Tamburrini, *Autonomia delle macchine e filosofia dell'intelligenza artificiale*, in *Rivista di filosofia*, 58, 2, 2017, p. 263 ss.

[51] A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019.

medico e paziente, ma anche in merito alla gestione del rischio da parte della struttura sanitaria che utilizzi sistemi algoritmici, atteso che il *software* di IA è un prodotto e, come tale ad esso si applica la relativa disciplina giuridica, nonché la ripartizione delle responsabilità tra il produttore dell'*hardware*, l'ideatore dell'applicazione, il programmatore del *software*, il responsabile del *training* e l'addetto alla sua manutenzione. È la cronaca a narrare il caso di un *robot* che ha provocato il decesso di un paziente in seguito alla somministrazione di una dose troppo elevata di radioterapia[52], così come in letteratura sono riportati i danni nel settore della chirurgia robotica statunitense, in particolare quelli cagionati dal *robot* "Da Vinci"[53], il più diffuso nel mondo, che ha causato più di cinquecento contenziosi legali nel solo anno 2021[54]. L'incertezza nella regolazione dei rapporti, in tali ambiti, è dovuta alla mancanza di norme che regolano le peculiarità del regime giuridico per i sistemi di IA, a partire dal loro "*status*" e, anzi, le normative vigenti nelle varie nazioni attribuiscono la responsabilità a volte al produttore, in altri casi al programmatore, ovvero all'utilizzatore di sistemi di IA. È di tutta evidenza che i giuristi sono chiamati a elaborare principi e prescrizioni puntuali al fine di governare l'IA. In tal senso, sono diverse le teorie sulla responsabilità dei sistemi di IA, che, da un lato, può essere assimilata alla responsabilità del produttore per danno da prodotto difettoso, disciplinata dalla Direttiva 85/374/CEE, e, d'altro canto, alla responsabilità contrattuale o extracontrattuale. Di conseguenza, in alcuni ordinamenti, la disciplina vigente si presta a risolvere i problemi dei danni causati dai sistemi di IA, mentre, in altri, si rende necessario un intervento riformatore del legislatore[55].

3.2 Le diverse teorie della dottrina sulla responsabilità dell'IA

Nel nostro ordinamento, l'art. 2043 del codice civile stabilisce che «qualunque fatto doloso o colposo, che cagiona ad altri un danno ingiusto, obbliga colui che ha

[52] C. Rossello, *Software difettoso e responsabilità da prodotto*, in *Economia e diritto del terziario*, 3, 1993, p. 975 ss.

[53] Jennifer Silvestrini v. University Healthcare System, L.C., d/b/a Tulane University Hospital and Clinic ("TUHC") e Intuitive Surgical, Inc., Case No. 11-2704, 2012; Starr and Griffith v. Intuitive Surgical, Inc., Case No. 1:14-cv-00225, 2014; Richard Pohly v. Intuitive Surgical, Inc., Case No. 15-cv-04113-MEJ, 2017.

[54] G. Guerra, *Diritto comparato e robotica: riflessioni sulla litigation americana in materia di chirurgia robotica*, in *Dir. inf.*, 2016, 2, p. 157 ss.; G. Guerra, *Profili di responsabilità del produttore di robot chirurgico nell'ordinamento americano*, in *Responsabilità medica. Diritto e pratica clinica*, 2, 2020, p.215 ss.

[55] U. Salanitro, *Intelligenza artificiale e responsabilità: la strategia della Commissione europea*, in *Rivista di diritto civile*, anno LXVI, 6, novembre-dicembre 2020.

commesso il fatto a risarcire il danno»[56]. Poiché tale precetto costituisce una formula generale dei fatti costitutivi la responsabilità aquiliana o extracontrattuale[57], non dovrebbe richiedere modifiche per poter essere adattato ai sistemi intelligenti, sebbene, con l'avanzare dell'industrializzazione, si sono moltiplicate le situazioni suscettibili di causare un danno, in riferimento alle quali l'imputazione per colpa non sembrava sufficiente per garantire il risarcimento del danneggiato[58]. Senonché, il legislatore del 1942 ha previsto una disciplina peculiare per alcuni casi, rispetto alla clausola generale dell'art. 2043 c. c. Si tratta degli articoli da 2047 a 2054 c. c., che disciplinano il danno cagionato dall'incapace, la responsabilità dei genitori, dei tutori, dei precettori e dei maestri d'arte, la rovina di edificio, la circolazione di veicoli, il danno cagionato da cose in custodia, l'uso strumentale dell'intelligenza umana altrui e di quella animale. In tali casi, è caratteristica la possibilità di una prova liberatoria a favore del soggetto chiamato a rispondere, su cui incombe l'onore della prova in giudizio, prima di una eventuale condanna al risarcimento del danno. In tale contesto, rileva lo *status* dell'agente. Difatti, ai sensi dell'art. 2046 c.c., «non risponde delle conseguenze dal fatto dannoso chi non aveva la capacità di intendere o di volere al momento in cui lo ha commesso, a meno che lo stato d'incapacità derivi da sua colpa». Di conseguenza, se si è chiamati a rispondere dei danni cagionati dalla propria condotta per avere o meno, al momento del compimento del fatto, (non la

[56] Corte di Cassazione, Sezione III civile, 4 febbraio 2014, n. 2422, in *CED Cassazione*, 2014, in cui si afferma che «gli elementi costitutivi dell'illecito aquiliano sono la condotta, l'elemento psicologico, il danno ingiusto e il nesso causale», pertanto, «ove il giudice ritenga insussistente uno qualsiasi di tali elementi, la domanda di risarcimento del danno va rigettata senza necessità di accertare la sussistenza degli altri».

[57] Per i profili penalistici, non potendo essere affrontati nella presente trattazione per ragioni di spazio, si rinvia, *ex multis*, a U. Pagallo, *Saggio sui robot e il diritto penale*, in Vinciguerra, Dassano (a cura di), *Scritti in memoria di Giuliano Marini*, Napoli, 2010; F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto Penale e uomo*, 29 settembre 2019; M. B. Magro, *Biorobotica, robotica e diritto penale*, in D. Provolo, S. Riondato, F. Yenisey (a cura di), *Genetics, Robotics, Law, Punishment*, Padova University Press, 2014; U. Pagallo, *The Laws of Robots. Crimes, Contracts and Torts*, Springer, 2013; P. M. Asaro, *Determinism, machine agency, and responsibility*, in *Politica & società*, 2014; S. Beck, *Google Cars, Software Agents, Autonomous Weapons Systems – New Challanges for Criminal Law*, in Hilgendorf-Seidel (eds.), *Robotics, Autonomics and the Law*, Baden-Baden, 2017; D. Lima, *Could AI Agents Be Held Criminally Liable? Artificial Intelligence and the Challanges for Criminal Law*, in *South Carolina Law Review*, 2018; S. Riondato, *Robot: talune implicazioni di diritto penale*, in Moro, Sarra (a cura di), *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, Milano, 2017; I. Salvadori, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *Rivista Italiana di Diritto e Procedura Penale*, 1, 2021; G. Hallevy, *The Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal*, 4, 2, 2010; C. Corridori, *Machina delinquere non potest*, in *Giustizia Insieme*, 19 maggio 2022; A. Cappellini, *Machina delinquere non potest? brevi appunti su intelligenza artificiale e responsabilità penale*", trascrizione della relazione orale tenuta alla giornata di studi "Il ragionamento giuridico nell'era dell'intelligenza artificiale", svoltasi presso l'università di Firenze il 15 novembre 2018.

[58] F. Galgano, *Trattato di diritto civile*, CEDAM, 2014, p. 129 ss.

capacità di agire, ma) la capacità di intendere o di volere, un soggetto ritenuto "incapace" dall'ordinamento può essere obbligato al risarcimento dei danni scaturiti dalla propria azione[59]. Parte della dottrina, allora, sostiene l'applicabilità all'IA degli istituti della responsabilità vicaria, e, in particolare, di quella del danno cagionato dall'incapace di intendere e di volere, di cui all'art. 2047 c. c. ("Danno cagionato dall'incapace"), che impone la riparazione del danno a chi è tenuto alla sorveglianza dell'incapace, in una sorta di responsabilità diretta del sorvegliante per *culpa in vigilando*[60], poiché l'unica ipotesi "pura" di responsabilità per fatto altrui sarebbe, invece, quella prevista nell'art. 2049 c. c. (responsabilità dei padroni e dei committenti derivante dalla condotta illecita tenuta dai propri domestici). Orbene, se ai soggetti ritenuti responsabili è offerta dall'ordinamento una prova liberatoria, dimostrando di non aver potuto impedire il fatto, è problematico stabilire chi potrebbe essere chiamato a sorvegliare un sistema di IA e a risarcire un eventuale danno causato, dal medesimo sistema, a terzi. In tal modo, al "proprietario" del sistema di IA, si dovrebbe attribuire, al momento dell'acquisto, il dovere di "controllo" dello stesso al fine di impedire che possa arrecare danni, impostazione, questa, che esonererebbe da responsabilità i produttori e i venditori dell'IA. In dottrina, allora, si scorge la possibilità di applicare in via analogica gli artt. 2047 e 2048 c. c., considerando il sistema di IA come "soggetto incapace" e coinvolgendo chi ne ha la responsabilità, tenendo in considerazione alcune caratteristiche quali: la natura del *robot*; i compiti che è in grado di svolgere; il tipo di controllo che l'uomo esercita su di esso; il modo in cui interagisce, il rapporto di utenza e il grado di prossimità del *robot* con l'essere umano; il contesto in cui è impiegato[61]. Altra parte della dottrina giudica contraddittorio, tuttavia, definire incapaci macchine dotate di intelligenza artificiale che notoriamente è superiore alla media[62]. Peraltro, gli artt. 2048 e 2049 c.c. sono norme tarate sulle specificità della natura umana dell'essere e dell'intelligenza di cui si risponde, e non sull'IA[63].

---

[59] A. Torrente, P. Schlesinger, *Manuale di diritto privato*, XXII ed., a cura di F. Anelli, C. Granelli, Giuffrè Francis Lefebvre, Milano, 2015, p. 916.

[60] F. Frigida, *Responsabilità del sorvegliante dell'incapace, dei genitori e tutori, dei padroni e Committenti*, in *La responsabilità civile*, Giuffrè Francis Lefebvre, 2018, p. 1788 ss.

[61] B. Bisol, A. Carnevale, F. Lucivero, *Diritti umani, valori e nuove tecnologie. Il caso dell'etica della robotica in Europa*, in *International Studies in Phenomenology and Philosophy*, 2014, p. 248 ss.

[62] F. Ferretti, *Intelligenza artificiale e responsabilità civile nel settore sanitario*, in *Actualidad Jurídica Iberoamericana*, 18, febrero 2023, pp. 1852-1885.

[63] U. Ruffolo, *Per i fondamenti di un diritto della robotica self-learning: dalla machinery produttiva all'auto driverless: verso una responsabilità da algoritmo?*, in Aa. Vv., *Intelligenza artificiale e responsabilità*, Giuffrè, Milano, 2017, p. 27.

Cosicché, parte della dottrina tedesca descrive i sistemi di IA come dei flussi di informazione talmente complessi da potersi vedere attribuita, a determinate condizioni, una sorta di "identità sociale", in una soggettività giuridica parziale che non impedirebbe il coinvolgimento dell'individuo nella responsabilità della macchina. Poiché la medesima non è in grado di autodeterminarsi, non si potrà affermare che essa prende decisioni in grado di ripercuotersi sulla sfera giuridica degli altri consociati, di conseguenza, si dovrebbe applicare una variante della responsabilità per fatto degli ausiliari[64]. Per queste e altre criticità della responsabilità extracontrattuale, altra parte della dottrina rappresenta la necessità risolutiva di riconoscere una vera e propria personalità giuridica ai sistemi intelligenti, con la creazione della "persona elettronica"[65], e, addirittura, la proposta di creare un pubblico registro delle personalità elettroniche – simile a quello delle imprese – contenente un numero di serie per ogni prodotto, che consenta, agli utilizzatori dello stesso, di avere tutte le informazioni rilevanti; ciò consentirebbe di configurare una diretta responsabilità del *robot* per i danni eventualmente occorsi per sua causa[66]. Secondo alcune ricerche degli esperti di robotica dell'Università di Pisa, la coscienza altro non è che una elevatissima capacità di calcolo di cui è dotato il cervello umano, dunque, nel momento in cui si arriverà a tale risultato, stimato dal 2029 in avanti, la macchina disporrà di una coscienza[67]. La "persona elettronica" costituirebbe un *tertium genus* di responsabilità, oltre a quello delle persone fisiche e a quello delle persone giuridiche. Nella Carta della Robotica del 16 febbraio 2017, il Parlamento UE ha raccomandato alla Commissione di emanare norme di diritto civile sulla robotica, invitando gli Stati membri, stante il vuoto normativo in merito all'IA, a utilizzare gli

---

[64] G. Teubner, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, trad. P. Femia, Edizioni Scientifiche Italiane, Napoli, 2019; E. Esposito, *Artificial Communication? The Production of Contingency by Algorithms*, in *Zeitschrift Für Soziologie*, 46, 4, 2017, p. 249 ss.

[65] B. J. Koops, M. Hildebrandt, D. O. Jaquet-Chiffelle, *Bridging the Accountability Gap: Rights for New Entities in the Information Society?*, in *Minnesota Journal of Law, Science & Technology*, 11, 2, 2010, p. 599 ss., originariamente proposto nel 1992, e ripreso successivamente in ambito civilistico, da L. B. Solum, *Legal Personhood for Artificial Intelligences*, in *NCL Rev.*, 1992, p. 1231 ss.; cfr., amplius, T. Allen, R. Widdison, *Can Computers Make Contracts?*, in *Harvard J. L. & Tech.*, 25, 1996, p. 25 ss.; S. Wettig, E. Zehendner, *A Legal Analysis of Human and Electronic Agents*, in *Artificial Intelligence and Law*, 12, 1-2, 2004, p. 111 ss.; G. Teubner, *Soggetti giuridici…, op. cit.*; R. van den Hoven van Genderen, *Do We Need New Legal Personhood in the Age of Robots and AI?*, in M. Corrales, M. Fenwick, N. Forgó (Eds.), *Robotics, AI and the Future of Law, Perspectives in Law, Business and Innovation (PLBI)*, Springer Publishers, 2018, p. 15 ss..

[66] C. Leroux, R. Labruto et a., *A Green Paper on Legal Issues in Robotics*, The European Robotics Coordination Action, décembre 2012.

[67] G. Taddei Elmi, F. Romano, *Il robot tra ius condendum e ius conditum*, in *Informatica e diritto*, XXV, 2, 2016, p. 137 ss.; G. Buttazzo, *Coscienza artificiale: missione impossibile*, in *Mondo Digitale*, 1.1, 2002, p. 16 ss.

strumenti tradizionali delle responsabilità oggettive o per fatto altrui (vicarie) o con la gestione del rischio. Il Parlamento ha altresì invitato la Commissione a valutare l'«istituzione nel lungo termine di uno status giuridico specifico per i *robot* in modo che almeno quelli 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 i terzi»[68]. Un argomento che avvalora la tesi della necessità della creazione di un *tertium genus* di responsabilità è costituito dal fatto che la costruzione di una "persona elettronica" rappresenterebbe l'unico regime di responsabilità che prescinde dal coinvolgimento di ulteriori soggetti, quali il proprietario, il produttore, il programmatore del sistema di IA, e, di conseguenza, riduce a zero il rischio che il danneggiato non sia "coperto", ciò che può accadere, invece, con l'applicazione analogica delle forme di responsabilità presenti nel codice civile[69]. In realtà, tuttavia, secondo parte della dottrina[70], la posizione espressa, nella Risoluzione del 2017, dal Parlamento UE sul riconoscimento della "persona elettronica", è stata fraintesa da chi ha avanzato l'equiparazione tra la soggettività umana e quella della macchina, poiché la stessa Risoluzione non avrebbe proposto il riconoscimento di una soggettività ontologica (simile a quella delle persone fisiche), bensì soltanto l'attribuzione di una soggettività "ascritticia", ossia attribuita o fittizia, sulla scia di quella attribuita dall'ordinamento italiano, per ragioni pratico funzionali, alle società e agli enti, vale a dire alle persone giuridiche. Il discrimine tra queste due forme di soggettività sarebbe la differenza tra "riconoscimento" e "attribuzione". Pertanto, una macchina non potrebbe mai essere chiamata a rispondere direttamente dei danni procurati, se non come persona giuridica ascritticia, dunque portatrice di una soggettività soltanto parziale, come d'altronde sostenuto da certa dottrina privatistica tedesca, pocanzi accennata[71]. Su

[68] Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)), punto n. 59, lettera f). Cfr. E. Burgio, L. De Simone, *Intelligenza Artificiale e responsabilità civile*, in *MediaLaws - Rivista di diritto dei media*, 15 aprile 2021, che richiama, a sua volta, L. Marini, I. Aprea, *Le guidelines on regulating robotics: una sfida per il diritto dell'Unione*, in *Osservatorio sul diritto della bioetica*, 5, 2015, p. 1295 ss.

[69] E. Palmerini, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, in *Osservatorio del diritto civile e commerciale*, 2, 2020, p. 464 ss.

[70] G. Taddei Elmi, A. Contaldo, *Intelligenza artificiale. Algoritmi giuridici. Ius condendum o "fantadiritto"?*, Pacini giuridica, 2020; G. Taddei Elmi, *Il Quid, il Quomodo e il Quid iuris dell'IA. Una riflessione a partire dal volume "Diritto e tecnologie informatiche"*, in *Rivista Italiana di Informatica e Diritto*, 2, 2021, DOI: 10.32091/RIID0047

[71] G. Teubner, *Soggetti giuridici…*, *op. cit.*; E. Esposito, Artificial Communication…, *op. cit.* Per altre critiche mosse all'eventuale creazione della "persona elettronica", cfr. A. Amidei, *Robotica Intelligente e Responsabilità: profili e*

una posizione di contestazione alla creazione di un *tertium genus* di personalità si pone un gruppo di esperti dell'IA e della robotica, laddove sostiene che sia errato creare uno «*status* giuridico di "persona elettronica" per *robot* "autonomi", "imprevedibili" e "autodidatti"» sulla base della (errata, a loro dire) affermazione che la responsabilità per danno sarebbe impossibile da provare[72]. Sotto il profilo etico e legale, peraltro, gli esperti sostengono che non è configurabile una personalità giuridica per un *robot*, né sul modello della persona fisica, poiché il *robot* sarebbe titolare di diritti umani (dignità, integrità, remunerazione, cittadinanza) e ciò sarebbe in contraddizione con la Carta dei diritti fondamentali dell'Ue e la Convenzione per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali. Su simili posizioni si attesta il Comitato economico e sociale europeo, che in un parere afferma di essere contrario a qualsiasi forma di *status* giuridico per i *robot* o l'IA[73], così come l'UNESCO, che trova contraddittorio definire "persone" i *robot* poiché non possiedono libertà di volontà, intenzionalità, autostima, coscienza e senso di identità personale[74]. In accordo con quanto affermato nella Carta della Robotica del 2017, parte della dottrina italiana ritiene applicabili ai sistemi di IA le attuali norme del codice civile sulla responsabilità oggettiva, come responsabilità contrapposta a quella per colpa, della persona che è in grado, in determinate circostanze, di ridurre al minimo i rischi e far fronte agli impatti negativi[75]. Poiché i *robot* sono equiparabili a qualsiasi altro prodotto tecnologico, può, per essi, valere la responsabilità per danno da prodotto difettoso o, al più, la responsabilità degli utilizzatori finali e/o dei

---

prospettive evolutive del quadro normativo europeo, in *Intelligenza Artificiale e Responsabilità*, a cura di U. Ruffolo, Milano, Giuffrè, 2017.

[72] AA.VV., Open letter to the European Commission Artificial Intelligence and Robotics, http://www.robotics-openletter.eu/

[73] European Economic and Social Committee, Artificial Intelligence - The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society (own-initiative opinion), 31 maggio 2017, https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence-consequences-artificial-intelligence-digital-single-market-production-consumption-employment-and

[74] UNESCO, Report of COMEST on robotics ethics, 2017, https://unesdoc.unesco.org/ark:/48223/pf0000253952

[75] L. Coppini, *Robotica e intelligenza artificiale: questioni di responsabilità civile*, in *Politica del diritto*, 4, 2018, p. 723 ss. Cfr., *amplius*, U. Ruffolo, *Intelligenza artificiale: il diritto, i diritti, l'etica*, Giuffrè, Milano, 2020; G. Capilli, *La responsabilità per la produzione di robot*, in G. Alpa (ed.), *La responsabilità del produttore*, Milano 2019, p. 625 ss.; G. Capilli, *Responsabilità e Robot*, in *La nuova giurisprudenza civile commentata*, 35, 2, 2019, p. 621 ss.; G. Capilli, *I criteri di interpretazione della responsabilità*, in G. Alpa (ed.), *Diritto e intelligenza artificiale*, p. 393 ss., p. 457 ss.; M. Infantino, *La responsabilità per danni algoritmici: prospettive europee-continentali*, in *Responsabilità civile e previdenza*, 84, 2019, 1762 ss.; M. Ratti, *Riflessioni in materia di responsabilità civile e danno cagionato da dispositivo intelligente alla luce dell'attuale scenario normativo*, in *Contratto e Impresa*, 3, 2020, p. 1174 ss.; A. Amidei, *Intelligenza artificiale e responsabilità da prodotto*, in *Intelligenza artificiale. Il diritto, i diritti, l'etica*, a cura di U. Ruffolo, Giuffrè, Milano, 2020, p. 125 ss.

proprietari di tali sistemi. Nel nostro ordinamento, le ipotesi di responsabilità oggettiva riguardano le attività pericolose, le cose in custodia, gli animali e i veicoli (previste nel codice civile), nonché, secondo le Direttive 374/1985 e 34/1999 recepite nel Codice del Consumo, i prodotti[76]. Tralasciando in questa sede, per motivi di spazio, il dibattito dottrinario[77] sull'applicabilità degli artt. 2054 ("Circolazione di veicoli"), 2050 ("Responsabilità per l'esercizio di attività pericolose") e 2051 c. c., a tenore del quale «ciascuno è responsabile del danno cagionato dalle cose che ha in custodia, salvo che provi il caso fortuito» – tutti casi in cui costituisce prova liberatoria quella riguardante l'adozione di tutte le misure idonee ad evitare il danno – emergono evidenti difficoltà volendo applicare tali disposizioni ai sistemi più autonomi, di per sé imprevedibili, con le conseguenti criticità relative all'opacità delle scelte algoritmiche e ai *bias* non facilmente riconoscibili che potrebbero essere presenti nei *dataset* e/o nei codici. Inoltre, la prova liberatoria nella responsabilità extracontrattuale si ha se l'autore del fatto illecito fornisce la prova dell'assenza di sua colpa, mentre nella responsabilità oggettiva è necessario provare che il danno è dovuto a un evento fortuito imprevedibile e inevitabile, orbene, l'onere probatorio del caso fortuito potrebbe costituire una *probatio diabolica*.

3.3 L'applicazione delle norme sulla responsabilità oggettiva

In dottrina, ci si è chiesti se sia appropriato attribuire in questi casi la responsabilità all'imprenditore o al produttore, a titolo di responsabilità oggettiva, cui si ricorre, generalmente, per garantire al meglio la posizione del danneggiato e per attribuire la responsabilità al professionista che si avvalga della tecnologia, a prescindere dalla colpa[78]. Eppure, il modello della responsabilità oggettiva è tradizionalmente riconosciuto come il più adeguato a rispondere alle esigenze di tutela del danneggiato, sorte con l'avvento della produzione di massa, sulla base del concetto di rischio di impresa[79], sempreché sussista un rischio costante e calcolabile connesso allo svolgimento dell'attività d'impresa. Mentre, ove i danni siano pertinenti ad un rischio non calcolabile a priori poiché occasionale, non dovrebbe trovare applicazione

[76] G. Taddei Elmi, *Il Quid, il Quomodo…, op. cit.*

[77] Cfr. M. Costanza, *L'Intelligenza Artificiale e gli stilemi della responsabilità civile, Dottrina e attualità giuridiche – Intelligenza Artificiale e diritto*, in *Giurisprudenza Italiana*, 2019, p. 1687 ss.; L. Coppini, *Robotica e intelligenza artificiale…, op. cit.*, p. 724 ss.; G. Sartor, *Gli agenti software e la disciplina giuridica degli strumenti cognitivi*, in *Il diritto dell'informazione e dell'informatica*, 1, 2003, p. 55 ss.

[78] C. M. Bianca, *La responsabilità*, in *Diritto civile*, V, III edizione, Giuffré Francis Lefebvre, Milano, 2021, p. 720 ss.

[79] P. Trimarchi, *Rischio e responsabilità oggettiva*, Giuffrè, Milano, 1961; G. Smorto, *Il criterio di imputazione della responsabilità civile. Colpa e responsabilità oggettiva in civil law e common law*, in *Europa e diritto privato*, 2008, p. 423 ss.

la responsabilità oggettiva[80]. L'alternativa alla responsabilità oggettiva, come proposta nella Carta della Robotica del 2017, è una responsabilità basata sulla gestione dei rischi, secondo cui «responsabile non sarebbe il soggetto negligente, in quanto responsabile a livello individuale, bensì colui che in determinate circostanze sia in grado di minimizzare i rischi e affrontare l'impatto negativo»[81]. Nell'ottica di *product liability*, tale soggetto risulta essere il produttore del prodotto difettoso, come definito nella Direttiva 85/374/CEE, recepita nel nostro ordinamento mediante il D.P.R. n. 224/1988, confluito, successivamente, nel D.lgs. n. 206/2005 ("Codice del Consumo"). È il (presunto) danneggiato a dover dimostrare la difettosità del prodotto (ai sensi dell'art. 117, comma 3: «elementi che rendano plausibile la mancanza della sicurezza che il danneggiato poteva legittimamente attendersi, senza necessariamente individuare lo specifico difetto tecnico»), il pregiudizio patito, nonché il nesso di causalità tra il danno occorso e il difetto (art. 4), circostanza, questa, complicata nell'ambito dei prodotti ad alta complessità tecnologica. Grava sul produttore, invece, l'onere di provare una delle ipotesi di esclusione di responsabilità di cui all'art. 118 del Codice del Consumo, fra cui quella dei «rischi da sviluppo», con l'inversione dell'onere probatorio[82]. Secondo l'art. 7 della Direttiva *de qua*, tuttavia, il produttore non è responsabile se prova, tra l'altro, «che lo stato delle conoscenze scientifiche e tecniche al momento in cui ha messo in circolazione il prodotto non permetteva di scoprire l'esistenza del difetto», evento difficile, ovviamente, laddove il sistema di IA agisca autonomamente rispetto a come è stato programmato. Di conseguenza, poiché se il produttore riesce a dimostrare l'inesistenza del difetto al momento della messa sul mercato del prodotto, il danneggiato resta privo di tutela, parte della dottrina sostiene la «necessità di immergere e considerare le regole di responsabilità civile nel più ampio quadro della responsabilizzazione», valutando positivamente il coinvolgimento del principio dell'*accountability*, che consiste nel dovere di dare conto del proprio comportamento[83]. La Corte di Giustizia UE ha

[80] S. Landini, *Assicurazione e responsabilità*, Giuffrè, Milano, 2004; A. D. Candian, *Responsabilità civile e assicurazione*, Egea, Milano, 1993.

[81] Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).

[82] E. Burgio, L. De Simone, *Intelligenza Artificiale e responsabilità civile*, in *MediaLaws - Rivista di diritto dei media*, 15 aprile 2021.

[83] G. Comandé, *Intelligenza artificiale e responsabilità tra liability e accountability: il carattere trasformativo dell'IA e il problema della responsabilità*, in *Analisi Giuridica dell'Economia*, Il Mulino, 1, 2019, p. 169 ss.; G. Comandé, *Responsabilità ed accountability nell'era dell'Intelligenza Artificiale*, in F. Di Ciommo, O. Troiano, *Giurisprudenza e autorità indipendenti nell'epoca del diritto liquido: studi in onore di Roberto Pardolesi*, La Tribuna, Roma, 2018, p. 1001 ss.; G. Finocchiaro, *Intelligenza artificiale e responsabilità*, in *Contratto e impresa*, 2020, p. 713 ss.

affermato che nella nozione di «prodotto» ai sensi dell'art. 2 della Direttiva, ossia ogni "bene mobile", rientra anche l'algoritmo che "guida" il sistema intelligente. Quindi, il danneggiato può chiamare in causa anche il programmatore del medesimo algoritmo, da cui ha avuto origine il comportamento dannoso – integrando una responsabilità solidale – ma non il *trainer* ovvero colui che ha impartito le istruzioni affinché la macchina sia in grado di auto-apprendere da un *dataset*[84]. In conclusione, la posizione maggioritaria intende optare per una responsabilità del produttore, seppur "rivisitata"[85], con la conseguenza che diventa cruciale la quantificazione del grado di minimizzazione dei rischi e degli impatti dei sistemi di IA, in un profilo di allocazione del rischio, poiché, a rilevare non è "chi" e "come" sbaglia, bensì la necessità di garantire un ristoro al danneggiato, ad esempio, mediante la costituzione di un fondo al quale attingere, a prescindere dall'individuazione delle modalità dell'incidente o dell'errore, come avviene nei casi di clonazione delle carte di credito[86]. Un tale sistema rassicurerebbe i potenziali utilizzatori, conferendo loro la garanzia di un risarcimento della somma sottratta. Un'elencazione *ex ante* delle attività "ad alto rischio", dunque, elaborata sulla base di standard e valutazioni d'impatto, non è tassativa e definitiva, ma è modificabile dopo l'immissione sul mercato del prodotto, sicché la responsabilità oggettiva troverebbe applicazione nelle sole ipotesi, contemplate, di volta in volta, nel catalogo di attività che hanno un margine di pericolosità inevitabile (nonostante l'osservanza di un diligente impegno), atteso che il rischio è alto e l'illecito potrebbe arrecare un danno ingiusto a un diritto fondamentale, mentre, in tutti gli altri casi, troverebbe applicazione una

[84] Corte Giust. UE, (Grande Sezione), 21 dicembre 2011, (domanda di pronuncia pregiudiziale proposta dal *Conseil d'État* — Francia) — *Centre hospitalier universitaire de, Besançon/Thomas Dutrueux, Caisse primaire d'assurance, maladie du Jura*, (Causa C-495/10), nel cui Dispositivo i giudici del Lussemburgo affermano che: «La responsabilità di un prestatore di servizi che utilizzi, nell'ambito di una prestazione di servizi quale l'erogazione di cure in ambiente ospedaliero, apparecchi o prodotti difettosi di cui non sia il produttore ai sensi delle disposizioni dell'art. 3 della direttiva del Consiglio 25 luglio 1985, 85/374/CEE, […] e causi in tal modo danni al destinatario della prestazione non rientra nell'ambito di applicazione di tale direttiva. Quest'ultima non osta, pertanto, a che uno Stato membro istituisca un regime, come quello di cui alla causa principale, che preveda la responsabilità di un simile prestatore per i danni in tal modo cagionati, anche in assenza di qualunque colpa imputabile al medesimo, a condizione, tuttavia, che sia fatta salva la facoltà per la vittima e/o per il suddetto prestatore di invocare la responsabilità del produttore in base alla citata direttiva, qualora risultino soddisfatte le condizioni previste dalla medesima».

[85] A. Amidei, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, in *Giurisprudenza italiana*, 7, 2019, p. 1657 ss.; A. Amidei, *Intelligenza artificiale e responsabilità…*, *op. cit.*; U. Salanitro, *Intelligenza artificiale e…*, *op. cit.*, p. 1246 ss.; G. Capilli, *Responsabilità e Robot*, in *La nuova giurisprudenza civile commentata*, 35, 2, 2019, p. 621 ss.; M. Ratti, *Riflessioni in materia…*, *op. cit.*, p. 1173 ss.

[86] G. Finocchiaro, *Intelligenza Artificiale e protezione dei dati* personali, in *Giurisprudenza Italiana*, Luglio 2019, p. 1670 ss.

responsabilità fondata sulla colpa[87]. Con riferimento, in particolare, all'allocazione delle responsabilità per le decisioni prese dai sistemi algoritmici non autonomi utilizzati in campo medico-sanitario, gli errori di programmazione possono essere attribuiti al programmatore o al venditore del prodotto, ai sensi della Direttiva 85/374/CE. Delle eventuali omissioni negli aggiornamenti del sistema, viceversa, ai sensi dell'art. 1218 c.c., possono essere chiamate a rendere conto le società incaricate della manutenzione, salvo che sia stata la struttura sanitaria a "dimenticare" di provvedere agli aggiornamenti, ricorrendo in tali casi l'applicabilità dell'art. 2051 del codice civile. Infine, laddove fosse l'operatore sanitario a fare un uso errato dell'IA (anche per averne soltanto interpretato in maniera fallace gli *output*) può vedersi applicata la responsabilità aquiliana. Nei casi di uso negligente o imperito del dispositivo opera, infatti, la specifica disciplina sulla responsabilità medica di cui all'art. 7 della legge n. 24/2017 ("Gelli-Bianco")[88], che imputa la responsabilità al medico, in via extracontrattuale, e, alla struttura sanitaria, quella da inadempimento contrattuale[89]. L'operatore sanitario è responsabile per l'impiego errato di tecnologie robotiche automatiche (ma non autonome), poiché esse sono considerate alla stessa stregua di un bisturi, lavorando, il *robot*, sotto la regia umana[90]. Appaiono evidenti, tuttavia, le residue criticità in merito sia all'onere della prova relativa al nesso causale tra la condotta del medico e il danno, che nel caso della responsabilità extracontrattuale è in capo al danneggiato dal sistema di IA e l'onere della prova "liberatoria" in capo alla struttura sanitaria, che non è chiaro se debba essere riferito all'ottemperanza delle "linee guida cliniche e scientifiche" o se debba essere esteso ad un uso dell'IA conforme a manuali d'uso o codici di condotta. Senza dimenticare che, per i casi in cui fosse invece chiamato alla prova liberatoria il produttore del sistema intelligente, egli, chiamato a dimostrare l'esistenza di una delle cause di "esclusione della responsabilità" (art. 118 del Codice del Consumo) – essendo irrilevante la prova di aver impiegato la diligenza richiesta nella fabbricazione o distribuzione del

[87] M. Rabitti, *Il danno prodotto…*, *op. cit.*

[88] Legge 8 marzo 2017, n. 24 "Disposizioni in materia di sicurezza delle cure e della persona assistita, nonché in materia di responsabilità professionale degli esercenti le professioni sanitarie".

[89] G. Alpa, *«Ars interpretandi» e responsabilità sanitaria nella nuova legge Bianco-Gelli*, in *Contr. impr.*, 2017, p. 728 ss.; R. De Matteis, *Le responsabilità in ambito sanitario. Il regime binario: dal modello teorico ai risvolti applicativi*, Giuffrè, Milano, 2017.

[90] C. Perlingieri, *Responsabilità civile e robotica medica*, in *Tecn. Dir.*, 2020, p. 171. 38 Cfr. anche Corte di Cassazione, 11 maggio 2009, n. 10743, in *Nuova giur. civ. comm.*, 2009, p. 1276, in cui il giudice nomofilattico ha affermato che "è pur sempre configurabile una responsabilità autonoma e diretta della struttura ospedaliera, ove il danno subito dal paziente risulti causalmente riconducibile ad un'inadempienza alle obbligazioni ad essa facenti carico".

prodotto – dovrebbe provare quasi l'impossibile, per tutta una serie di motivi legati alle caratteristiche intrinseche dell'IA. Come accennato, la responsabilità del produttore è disciplinata dalla Direttiva 85/374/CEE, recepita in Italia agli artt. 114-127 del Codice del Consumo, in un regime di *strict liability*, in virtù del quale il danneggiato ha l'onere di allegare l'inadempimento e di provare il danno subito, nonchè il nesso causale tra inadempimento e danno (art. 120)[91], come ormai da tempo richiede la giurisprudenza[92]. Per ovviare a questi e altri problemi applicativi, la Commissione europea, congiuntamente alla proposta legislativa volta ad adeguare le norme in materia di responsabilità civile extracontrattuale all'IA, in data 28 settembre 2022, ha presentato un ulteriore testo legislativo, al fine di modificare molteplici aspetti della disciplina attualmente in vigore in materia di responsabilità per danno da prodotti difettosi[93]. La proposta tende a includere nella definizione di "prodotto" anche l'elettricità, il software e i file per la fabbricazione digitale. Inoltre, l'onere della prova resterebbe a carico del danneggiato; tuttavia, nella proposta sarebbero introdotte delle presunzioni in merito al carattere difettoso dei prodotti, al nesso di causalità tra difetto e danno o ad entrambi gli elementi. Soprattutto, la proposta mira a far sì che il consumatore abbia la garanzia che esista nell'Unione, in ogni momento, un operatore economico (rappresentante) contro il quale possa essere presentata una domanda di risarcimento. Infine, secondo la proposta, in discussione presso il Parlamento Europeo, la responsabilità non dovrebbe essere esclusa, né limitata per legge o disposizione contrattuale, non consentendo, dunque, agli Stati membri, di fissare soglie minime o massimali finanziari per il risarcimento. Appare, dunque, evidente che qualificare l'algoritmo quale "componente" del prodotto finale, permetterebbe di riconoscere anche in capo al suo ideatore, se diverso dagli altri produttori, una responsabilità extracontrattuale autonoma, come, d'altronde, richiedeva da tempo parte della dottrina[94]. Sotto tali profili, il danneggiato sarebbe

[91] F. Ferretti, *Intelligenza artificiale…*, *op. cit.*

[92] Corte di Cassazione, 11 novembre 2020, n. 25288; 11 novembre 2019, n. 28991; 26 luglio 2017, n. 18392, tutte in www.dejure.it: "Ove sia dedotta una responsabilità contrattuale della struttura sanitaria per l'inesatto adempimento della prestazione sanitaria, è onere del danneggiato provare il nesso di causalità fra l'aggravamento della situazione patologica (o insorgenza di nuove patologia per effetto dell'intervento) e l'azione o l'omissione dei sanitari, mentre è onere della parte debitrice provare che una causa imprevedibile ed inevitabile ha reso impossibile l'esatta esecuzione della prestazione (fattispecie relativa alla controversia promossa da una donna al fine di veder riconosciuto il risarcimento del danno per la morte del marito avvenuta nel corso di un intervento chirurgico)".

[93] Commissione UE, COM (2022) 495 final del 28.09.2022, Proposta di Direttiva del Parlamento e del Consiglio sulla responsabilità per danno da prodotti difettosi.

[94] U. Ruffolo, *Intelligenza Artificiale e Responsabilità*, *op. cit.*, p. 1689.

avvantaggiato, poiché vedrebbe ampliarsi la platea di soggetti chiamati a risarcire il danno subìto. Senza dimenticare la specifica presunzione circa la difettosità del prodotto per la violazione degli standard tecnici (art. 9, comma 2, lett. b), voluta dalla Commissione, in un'ottica di rafforzamento della relazione tra regole di sicurezza e di responsabilità, laddove ha affermato che «le caratteristiche dell'intelligenza artificiale non precludono il diritto a pretendere che i prodotti siano sicuri, che si tratti di tosaerba automatici o di robot chirurgici", sebbene un prodotto conforme non sia, di converso, automaticamente, privo di difetti»[95]. Ciononostante, parte della dottrina ha negato che la responsabilità del sanitario rientri nell'ambito di applicazione della direttiva 85/374 CEE, con relativa esclusione della disciplina consumieristica in favore del paziente danneggiato, tesi la cui *ratio* è da ricercarsi nell'assenza, in capo alla struttura sanitaria, di qualsiasi fine di profitto, da cui deriva l'impossibilità di qualificarla come "imprenditore" o "professionista"[96].

3.4 I principi della Proposta di Direttiva UE sulla responsabilità extracontrattuale dell'IA

Agli interrogativi della dottrina in merito alla responsabilità civile extracontrattuale dell'IA sembra aver dato parziali risposte la Proposta di Direttiva presentata dalla Commissione Europea, nel settembre 2022[97], laddove ha inteso risolvere il problema del nesso di causalità dell'evento dannoso mediante il meccanismo della presunzione relativa, mantenendo come il sistema dell'*AI Act* del "doppio binario" basato sul rischio dei sistemi di IA. Nello spazio europeo, infatti, oltre al Regolamento in via di adozione, verrà in rilievo (entro due anni) anche tale Direttiva, nata dall'esigenza di «migliorare il funzionamento del mercato interno stabilendo requisiti uniformi per determinati aspetti della responsabilità civile extracontrattuale per danni causati con il coinvolgimento di sistemi di IA»[98]. Nella Relazione illustrativa si afferma, infatti, che «le norme nazionali vigenti in materia di responsabilità, in particolare per colpa, non sono adatte a gestire le azioni di responsabilità per danni causati da prodotti e servizi basati sull'IA», di conseguenza «se un danneggiato presenta una domanda di risarcimento, gli organi giurisdizionali nazionali, confrontati alle caratteristiche

[95] F. Ferretti, *Intelligenza artificiale…*, *op. cit.*
[96] F. Ferretti, *Ibidem*.
[97] Proposta di DIRETTIVA DEL PARLAMENTO EUROPEO E DEL CONSIGLIO relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale (direttiva sulla responsabilità da intelligenza artificiale), https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX%3A52022PC0496.
[98] *Ibidem*, Relazione illustrativa.

specifiche dell'IA, possono adeguare le modalità di applicazione delle norme vigenti in base ai singoli casi per giungere a un risultato equo per il danneggiato. Ciò provocherà incertezza giuridica. Le imprese avranno difficoltà a prevedere le modalità di applicazione delle norme vigenti in materia di responsabilità e, di conseguenza, a valutare la loro esposizione alla responsabilità e a stipulare la relativa copertura assicurativa»[99]. A titolo esemplificativo, si pensi ai casi di incidenti in cui siano coinvolte auto autonome, eventi nei quali la responsabilità dei danni conseguenti potrebbe essere addebitata (in concorso o singolarmente) al conducente, a chi ha progettato il sistema di guida autonoma, a chi ha prodotto il software che guida i sensori, al produttore dei sensori, alla casa produttrice dell'auto, a eventuali soggetti che hanno messo a disposizione i dati di *training* dell'algoritmo, nonché, in caso di attacco informatico, ai cyberattaccanti. La Direttiva è strutturata su un approccio a due fasi. Nella prima fase, l'obiettivo è quello di coordinare i regimi nazionali in materia di responsabilità civile per danni causati dall'IA, mediante un "alleggerimento" dell'onere della prova per i danneggiati e, in particolare, della prova del nesso causale tra colpa del convenuto ed evento consistente in un'azione od omissione derivante dall'impiego di sistemi di IA. Nella seconda fase, verranno valutate le misure adottate in una sorta di monitoraggio che dovrà tener conto soprattutto, da un lato, degli sviluppi tecnologici, normativi e giurisprudenziali intervenuti, e, d'altro canto, della necessità di completare l'armonizzazione degli ordinamenti nazionali in materia di responsabilità oggettiva e copertura assicurativa. In generale, è emersa la volontà, del legislatore europeo, di incanalare lo sviluppo tecnologico verso un'intelligenza artificiale antropocentrica, evitando, però, di frenare il mercato sottostante per non perdere terreno nei confronti di altri Paesi[100]. La Proposta *de qua* stabilisce che le domande di risarcimento del danno possono essere presentate, oltre che dalla persona danneggiata, anche da persone che sono subentrate o si sono surrogate nei diritti del danneggiato[101]. La surrogazione, che in

[99] *Ibidem*, Relazione illustrativa.

[100] Nell'Indagine sulla diffusione delle tecnologie basate sull'intelligenza artificiale tra le imprese europee, Ipsos 2020, relazione finale, pag. 58, https://op.europa.eu/en/publication-detail/-/publication/f089bbae-f0b0-11ea-991b-01aa75ed71a1, la responsabilità è emersa quale uno dei tre ostacoli principali all'uso dell'IA da parte delle imprese europee, venendo citata come l'ostacolo esterno più rilevante (43 %) per le imprese che intendono adottare l'IA, ma che non lo hanno ancora fatto.

[101] Proposta di DIRETTIVA DEL PARLAMENTO EUROPEO E DEL CONSIGLIO relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale (direttiva sulla responsabilità da intelligenza artificiale), https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX%3A52022PC0496, articolo 2, punto 6, lettera b).

tali casi riguarderebbe anche gli eredi di un danneggiato deceduto, è l'assunzione da parte di un terzo (come una compagnia di assicurazioni) del diritto giuridico di un'altra parte di riscuotere un debito o un risarcimento del danno. Inoltre, un'azione di risarcimento del danno può essere intentata anche da chi agisce per conto di uno o più danneggiati conformemente al diritto dell'Unione o nazionale[102], ciò per offrire maggiori possibilità alle persone danneggiate da un sistema di IA di far esaminare le loro domande da un organo giurisdizionale, anche nei casi in cui le azioni individuali possono sembrare troppo costose o troppo onerose da intentare, o in cui le azioni congiunte possono comportare un beneficio di scala. Ancora, la Proposta mira a fornire alle persone che chiedono il risarcimento di danni causati da sistemi di IA ad alto rischio mezzi efficaci per identificare le persone potenzialmente responsabili e le prove pertinenti per una domanda di risarcimento, prevedendo che un organo giurisdizionale possa ordinare la divulgazione di elementi di prova rilevanti in relazione a specifici sistemi di IA ad alto rischio che si sospetta abbiano cagionato danni[103]. Le richieste di elementi di prova sono rivolte al fornitore di un sistema di IA, a una persona soggetta agli obblighi del fornitore, della legge sull'IA o a un utente a norma della legge sull'IA. Le richieste dovrebbero essere suffragate da fatti e prove sufficienti a dimostrare la plausibilità della domanda di risarcimento del danno in questione e gli elementi di prova richiesti dovrebbero essere a disposizione dei destinatari. Le richieste non possono essere rivolte a parti sulle quali non incombono obblighi a norma della legge sull'IA e le quali non hanno pertanto accesso agli elementi di prova. Con riferimento ai danni causati dai sistemi di IA, la direttiva mira a fornire una base efficace per le domande di risarcimento in relazione alla colpa consistente nella non conformità a un obbligo di diligenza a norma del diritto dell'Unione o nazionale. Poiché è di difficile individuazione il nesso causale tra la non conformità e l'*output* prodotto dal sistema di IA o la mancata produzione di un *output* da parte del sistema di IA che ha cagionato il danno in questione, nella Proposta oggetto della presente analisi è stata stabilita una presunzione relativa mirata di causalità in relazione a tale nesso di causalità[104]. La colpa del convenuto deve essere dimostrata dall'attore conformemente alle norme dell'Unione o nazionali applicabili e può essere accertata, ad esempio, per non conformità a un obbligo di diligenza a norma della legge sull'IA o di altre normative stabilite a livello di Unione,

[102] *Ibidem*, articolo 2, punto 6, lettera c).
[103] *Ibidem*, articolo 3, par. 1.
[104] *Ibidem*, articolo 4, par. 1.

come quelle che disciplinano l'uso del monitoraggio e del processo decisionale automatizzati per il lavoro mediante piattaforme digitali o quelle che disciplinano il funzionamento di aeromobili senza equipaggio. Nei paragrafi 2 e 3 dell'articolo 4, la Proposta legislativa segna una distinzione tra le azioni intentate nei confronti del fornitore di un sistema di IA ad alto rischio o di una persona soggetta agli obblighi del fornitore a norma della legge sull'IA, da un lato, e le azioni intentate nei confronti dell'utente di tali sistemi, dall'altro, operando, in tali casi, il Reg. UE sull'IA. Ove vi fossero simili domande di risarcimento, l'osservanza da parte dei convenuti degli obblighi di cui al paragrafo 2 dell'art. 4, deve essere valutata anche alla luce del sistema di gestione dei rischi e dei suoi risultati, vale a dire le misure di gestione dei rischi, a norma della legge sull'IA. In caso di sistemi di IA ad alto rischio, quali definiti dal Reg. UE sull'IA, la Proposta stabilisce un'eccezione alla presunzione di causalità qualora il convenuto dimostri che l'attore può ragionevolmente accedere a elementi di prova e competenze sufficienti per dimostrare il nesso causale[105]. Nel caso, invece, di sistemi di IA non ad alto rischio, la Proposta stabilisce una condizione per l'applicabilità della presunzione di causalità, in base alla quale quest'ultima è soggetta alla condizione che l'organo giurisdizionale stabilisca che è eccessivamente difficile per l'attore dimostrare l'esistenza del nesso causale[106]. Tali difficoltà, si afferma nella Relazione illustrativa, «devono essere valutate alla luce delle caratteristiche di determinati sistemi di IA, come l'autonomia e l'opacità, che nella pratica rendono estremamente difficile la spiegazione del funzionamento interno del sistema di IA e pregiudicano la capacità dell'attore di dimostrare l'esistenza del nesso di causalità tra la colpa del convenuto e l'output del sistema di IA». Il convenuto ha ovviamente il diritto di confutare la presunzione di causalità fondata sull'articolo 4, paragrafo 1.

4. IA in medicina

La mancanza di una chiara configurazione delle responsabilità nei casi accertati di danno conseguente ad utilizzi di strumenti di IA è ancor più sentita nel contesto della medicina e la disciplina accennata potrebbe non essere completamente funzionale in tale ambito, come affermato in dottrina[107]. Con il termine "*e-health*" si

[105] *Ibidem*, articolo 4, par. 4.

[106] *Ibidem*, articolo 4, par. 5.

[107] F. Ferretti, *Intelligenza artificiale…*, *op. cit.*: «Non è chiaro, però, quale sarà l'impatto dell'intervento europeo in materia civilistica, soprattutto in relazione al sottosistema della responsabilità medica: la direttiva, una volta

definiscono attività come operazioni chirurgiche "a distanza", telemedicina, diagnosi e prevenzione, robotica riabilitativa e di assistenza. L'utilità dell'uso dell'IA in campo sanitario si manifesta, infatti, non soltanto nella razionalizzazione dell'uso delle risorse, ma anche nella possibilità di "personalizzare" i percorsi assistenziali, almeno per le patologie croniche, con conseguente miglioramento della qualità della vita del paziente. Inoltre, come osservato in dottrina, le tecnologie basate sull'IA hanno, da un lato, il potenziale di trasferire diverse procedure diagnostiche, tipiche delle rispettive branche specialistiche, nel *setting* delle cure primarie, democratizzando l'accesso alle cure e aprendo scenari innovativi nell'ambito delle prestazioni territoriali, e, sotto un diverso profilo, è possibile anche, grazie alla digitalizzazione dei dati, una gestione intelligente ed efficace del paziente cronico, con l'introduzione di nuove potenzialità nella ricerca clinica, nella medicina di iniziativa e nella medicina di personalizzata[108]. In Italia, il legislatore ha attribuito al Ministero della salute la facoltà di perseguire la razionalizzazione della spesa a mezzo di modelli predittivi[109]. La competenza nell'eventuale normazione di modelli di medicina predittiva spetta alle Regioni alle Province autonome. Orbene, in tali ambiti, non è semplice addebitare a un unico soggetto la responsabilità dell'evento dannoso, in particolare quella penale, che, come noto, secondo la Costituzione, è «personale»[110], considerata la molteplicità dei soggetti agenti oltre al medico (struttura sanitaria proprietaria dell'apparecchiatura, produttore del software, programmatore,

---

recepita, potrebbe sostituire le regole esistenti laddove si faccia uso di AI in sanità, oppure potrebbe affiancarsi ad esse, creando un "doppio binario", lasciando la scelta al paziente sull'esperibilità del rimedio; oppure ancora, potrebbe inserirsi nel sistema della responsabilità apportando una modifica al solo aspetto dell'accertamento del nesso causale nella correlazione tra fatto ed evento. Inoltre, il passaggio dalla responsabilità oggettiva (prevista nella risoluzione del 2020 per i sistemi di AI ad alto rischio) a questa ipotesi "sui generis" di responsabilità parzialmente presunta, che rimane ancorata alla valutazione della colpa e del canone della diligenza, rischia di mantenere nel sistema un elemento di incertezza poco funzionale alle esigenze di stabilità e prevedibilità manifestate dagli operatori del mercato. Secondo una diversa prospettiva poi, il regime di liability potrebbe non essere un rimedio sempre adeguato, soprattutto in relazione a danni provocati da errori fisiologici, ineliminabili dai device medici, ma calcolabili a priori, che non sembra opportuno né lasciare definitivamente sulla vittima, né traslare sull'utilizzatore o sul produttore, e che andrebbero invece gestiti attraverso una socializzazione dei costi garantita da fondi sovvenzionati dai soggetti coinvolti (cd. compensation fund)».

[108] A. Malva, V. Zurlo, *La medicina nell'era dell'Intelligenza Artificiale: applicazioni in Medicina Generale*, in *Rivista Società Italiana di Medicina Generale*, vol. 26, 4, 2019.

[109] Art. 7, comma 1, D.-l. 34/2020: «il Ministero può trattare […] ai sensi dell'articolo 2-*sexies*, comma 2, lettera v), del decreto legislativo 30 giugno 2003, n. 196 e nel rispetto del Regolamento UE 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016, dati personali, anche relativi alla salute degli assistiti, raccolti nei sistemi informativi del Servizio sanitario nazionale, nonché dati reddituali riferiti all'interessato e al suo nucleo familiare per lo sviluppo di metodologie predittive dell'evoluzione del fabbisogno di salute della popolazione, secondo le modalità di cui al decreto del Ministro della salute 7 dicembre 2016, n. 262».

[110] Art. 27, Costituzione.

sviluppatori, società addette alla verifica e manutenzione tecnica). Dal punto di vista normativo, infatti, tali dispositivi rientrano quasi sempre nella definizione di "dispositivi medici" – ossia «qualunque strumento, apparecchio, apparecchiatura, software, impianto, reagente, materiale o altro articolo destinato dal fabbricante a essere impiegato sull'uomo» – con applicazione della relativa disciplina, di recente innovata, a livello europeo, con il Reg. UE n. 745/2017, in vigore dal 26 maggio 2021, volto a garantire il buon funzionamento del mercato interno dei "device medicali", un livello elevato di protezione della salute dei pazienti, nonché il rispetto degli standard di qualità e sicurezza[111]. Inoltre, con la diffusione di sistemi di *computer vision*, di sonde e braccia meccaniche, le operazioni chirurgiche vengono sempre di più automatizzate, tanto che l'addestramento all'uso di sistemi di chirurgia robotica è, ormai, uno *step* necessario nella formazione medica specialistica, in particolare in nefrologia e urologia[112]. Le più moderne tecnologie di robotizzazione dell'intervento prevedono il suo svolgimento da parte del sistema di IA, che ha livelli variabili di autonomia, sotto il controllo, o l'integrale direzione, del medico, attraverso un monitor. Ad esempio, il *Da Vinci*, sistema robotico più diffuso al mondo, consente al chirurgo, fisicamente lontano dal campo operatorio e seduto a una postazione dotata di monitor e comandi, di muovere le braccia del *robot*, collegati agli strumenti endoscopici, che vengono introdotti attraverso piccole incisioni. Il campo operatorio è proiettato tridimensionalmente, con immagini ferme e ad altissima risoluzione[113]. Vi è, infine, un altro ambito in cui le criticità della disciplina legale dei sistemi di IA, quello della medicina "predittiva" (o "di iniziativa"), un modello volto a un superamento della "medicina d'attesa", che, alla luce delle informazioni genetiche di un soggetto sano, elabora una previsione del rischio di contrarre, nell'arco dell'esistenza, una determinata patologia. Grazie ad algoritmi "predittivi", che operano su grandi quantità di dati raccolti da *smartwatch*, bilance impedenziometriche e altri simili strumenti di allarme (peso, percentuale di massa grassa e magra, battito cardiaco, pressione, attività fisica svolta, ecc.) l'IA riesce ormai a prevedere possibili evoluzioni cliniche. Il ML è considerato promettente nell'ambito della pratica clinica, come supporto all'attività decisionale dei medici; basti pensare alla generazione di una stima del rischio di un paziente per uno specifico *outcome*,

[111] F. Ferretti, *Intelligenza artificiale…*, *op. cit.*

[112] T. Davenport, R. Kalakota, *The potential for artificial intelligence in healthcare*, in *Future Healthc J.*, 6, 2019, p. 94 ss.

[113] T.O. Jefferson et al., *Chirurgia robotica*, Roma, marzo 2017, finanziato da Agenzia nazionale per i servizi sanitari regionali (Agenas) e Ministero della Salute.

aiutando il medico a impostare terapie più consapevoli[114], o della *precision medicine*, per identificare i pazienti che potrebbero rispondere meglio ad una determinata terapia, indirizzando il medico verso la strategia terapeutica migliore[115]. Secondo alcuni recenti studi, tramite ML è possibile identificare nuovi predittori di rischio, migliorando l'accuratezza diagnostica quando paragonati a metodi tradizionali[116]. In oncologia, neurologia e cardiologia tali approcci sono ormai diffusi[117]. In tale contesto, per dirla con le parole di autorevole dottrina, "*Quali responsabilità allora, se la macchina sbaglia, ed il medico che la usa avalla l'errore?*"[118], ovvero se il sanitario scegliesse di suggerire un intervento chirurgico sulla base di un rischio segnalato dall'IA, poi non concretizzatosi. Per i casi in cui, viceversa, il medico dovesse disattendere la segnalazione dell'algoritmo, rivelatasi, invece, successivamente, fondata, si fa sempre più strada la proposta di configurare quell'accennato *tertium genus* di personalità giuridica, ossia la "personalità elettronica" della macchina. Il carattere "trasformativo" dell'IA, ovvero la sua capacità di generare profondi e dirompenti cambiamenti, è connesso ai suoi gradi di autonomia[119]. Gli esempi che possono essere annoverati sono molteplici, basti pensare alle microcapsule robotiche impiegate a fini terapeutici[120], oppure a un *software* che valuti la densità del seno durante l'esame mammografico, quale utile indizio di patologie ai fini diagnostici. Vi sono, altresì, meccanismi di ragionamento artificiale in grado di produrre una medicina predittiva, preventiva, personalizzata e partecipativa nella gestione del diabete[121], grazie al *deep learning* supervisionato con reti neurali[122] convoluzionali, che imitano il funzionamento del cervello umano e sono in grado di imparare a svolgere in modo automatico compiti tipicamente affidati alla sola

[114] E. Peterson, *Machine Learning, Predictive Analytics, and Clinical Practice: Can the Past Inform the Present?*, in *JAMA network*, 22(23), 2019, pp. 2283-2284.

[115] N.L. Guthrie, *Achieving Rapid Blood Pressure Control With Digital Therapeutics: Retrospective Cohort and Machine Learning Study*, in *JMIR Cardio*, 3, 2019.

[116] O. Colpani, *Machine Learning: la capacità di prevedere applicata alla ricerca e alla pratica clinica*, in *Giornale Italiano di Farmacoeconomia e Farmacoutilizzazione*, 11 (4), 2019, p. 5-11, in cui si rinvia a R.C. Deo, *Machine Learning in Medicine*, in *Circulation*, 2015, 132, pp. 1920-1930; E. Peterson, *Machine Learning, Predictive Analytics, and Clinical Practice: Can the Past Inform the Present?*, JAMA, 2019.

[117] V.J. Mar & H.P. Soyer, *Artificial Intelligence for Melanoma Diagnosis: How Can We Deliver on the Promise?*, in *ANN. ONCOL.*, 2018, p. 1628; F. Jiang et al., *Artificial Intelligence in Healthcare: Past, Present and Future*, in *STROKEVASC. NEUROL.*, 2017, p. 243 ss.; T. Davenport, R. Kalakota, *The potential for…*, *op. cit.*, pp. 94-98.

[118] U. Ruffolo, *Intelligenza artificiale, machine learning e responsabilità da algoritmo*, in *Giur. it.*, 2019, p. 1691 ss.

[119] G. Comandé, *Intelligenza artificiale…*, *op. cit.*

[120] A. Sieg, K. Friedrich, U. Sieg, *Is PillCam Colon capsule endoscopy ready for colorectal cancer screening? A prospective feasibility study in a community gastroenterology practice*, in *Am. J. Gastr.*, 2009, pp. 848-854.

[121] E. Samer, *Artificial Intelligence: The Future for Diabetes Care*, in *Am J Med*, 133, p. 895-900.

[122] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, in *Mit Press*, Cambridge, 2016, p. 8.

percezione visiva di operatori sanitari esperti. A differenza di altre applicazioni di IA che mettono semplicemente a confronto i dati raccolti – si pensi alla mera comparazione di due lastre radiologiche – in questi casi, la macchina compone i dati inseriti dall'umano e, mediante un'analisi e una valutazione accurata, riesce a esprimere, attraverso un procedimento decisionale autonomo (le reti neurali), un giudizio frutto di un ragionamento, che non richiede l'interpretazione dell'operatore. L'apprendimento profondo, una *species* più avanzata del *machine learning*, mediante le reti neurali permette all'IA di apprendere in modo automatico e non supervisionato da dati casuali, non precedentemente strutturati e catalogati. La problematica principale è costituita dalla non codificabilità, in numerosi casi, dell'*output* della macchina, poiché, a differenza del ragionamento umano, il modello predittivo non è abbastanza trasparente, tanto che si parla di opacità dell'algoritmo e le reti neurali sono definite modelli a scatola nera (*black box*[123]). In dottrina si è affermato come, in quanto tecnologia sanitaria, l'IA applicata alla ricerca, medicina, sanità, necessita di essere sviluppata attraverso una fase preclinica, clinica pilota e clinica allargata, nonché di essere sottoposta ad una formale valutazione (*Health Technology Assestment*), comprensiva delle sue implicazioni etiche, legali e sociali[124]. Difatti, se, da un lato, l'avvento delle reti neurali, e in particolare del *deep learning*, ha rappresentato una rivoluzione per il settore come avvenuto soltanto in pochi altri contesti[125], poiché ha permesso lo sviluppo di accurati sistemi valutativi e predittivi[126], d'altro canto, esso ha anche posto numerose questioni etiche. Come osservato in dottrina, l'applicazione del ML alla pratica clinica richiede la messa a punto di un percorso di approvazione che si adatti alle particolari caratteristiche dei software *ML-based*[127]. Ad esempio, in termini di qualità dell'algoritmo, come noto, tutto deriva dalla *quality* dei dati utilizzati nella fase di *training*. Si parla di *fairness*, ad

[123] S. Doncieux, J. Mouret, *Beyond black-box optimization: a review of selective pressures for evolutionary robotics*, in *Evolutionary Intelligence*, 7, 2014, p. 71 ss.; S. Beck, *Google cars, software agents, autonomous weapons systems – New challenges for criminal law?*, in E. Hilgendorf, U. Seidel (eds.), *Robotics, Autonomics, and the Law*, Baden-Baden, 2017, p. 227 ss.; G. Gaudio, *Algorithmic management, poteri datoriali e oneri della prova: alla ricerca della verità materiale che si cela dietro* l'algoritmo, in *Labour&Law Issue*, 6, 2, 2020; M. Busuioc, *Accountable Artificial Intelligence: Holding Algorithms to Account*, in *Public Administration Review*, 81, 5, 2020, p. 825 ss.

[124] A. Malva, V. Zurlo, *La medicina nell'era…*, *op. cit.*

[125] T. Davenport, R. Kalakota, *The potential for…*, *op. cit.*, pp. 94-98; E. Topol, *Deep medicine. How artificial intelligence can make healthcare human again*, New York, 2019; V. Kaul, S. Enslin, Seth A. Gross, *History of artificial intelligence in medicine*, in *Gastrointestinal Endoscopy*, 92, 4, 2020, pp. 807-812.

[126] S. F. Weng, J. Reps, J. Kai, J.M. Garibaldi, N. Qureshi, *Can machine-learning improve cardiovascular risk prediction using routine clinical data?*, in *PLOS ONE*, 12, 4, 2017.

[127] O. Colpani, *Machine Learning…*, *op. cit*., p. 5-11.

esempio, per il caso di un algoritmo generato da dati ottenuti solo da soggetti caucasici, avrebbe capacità predittive ridotte se applicato a soggetti asiatici o africani[128]. È appurato che i modelli di ML funzionano meglio quanto più è ampio il *training set*. Evitare, o almeno mitigare, siffatti *bias* statistici è possibile preferendo dati provenienti da studi randomizzati, invece che da studi osservazionali, dati non dipendenti dal giudizio del medico, o utilizzando strumenti volti a valutare il rischio di *bias* nell'algoritmo[129]. Ciò va bilanciato, tuttavia, con la normativa privacy, gli atti di indirizzo del Garante[130] e con l'esigenza di accedere a dati vasti ed eterogenei[131]. Non mancano, dunque, i tentativi di progettare sistemi di IA trasparenti (*clear box*), con l'obiettivo di rendere espliciti i modelli e i ragionamenti sottesi alle decisioni algoritmiche, consentendo al sanitario di orientarne il ragionamento, e, in particolare, sistemi "nativamente spiegabili", abbracciando soluzioni di IA trasparente mediante l'adozione di uno strumento di ML a regole esplicite, la *Logic Learning Machine* (LLM)[132]. Tale tipo di ML genera immediatamente e contestualmente alla creazione del modello, tutta una serie di informazioni esplicite, in linguaggio naturale e non sotto forma di equazioni matematiche, che permettono di capire quali siano le dinamiche che influenzano il fenomeno analizzato[133]. Di conseguenza, si avranno risposte affidabili alle seguenti domande[134]:

1. 'How the model thinks?' (internal reasoning). Ovvero: quali sono le logiche sottostanti al funzionamento del modello, cosa ha appreso e in base a cosa farà le previsioni.
2. 'What will be' (predictive analytics). Ovvero: cosa succederà?
3. 'Why did it happen?' (new finding – comprensione delle dinamiche del fenomeno). Ovvero: perché è successo ciò che è successo? Quali sono i motivi per cui probabilmente succederà ancora?

---

[128] CJ. Kelly et al., *Key challenges for delivering clinical impact with artificial intelligence*, in *BMC Med.*, 2019, 17, 195.
[129] R. B. Parikh, S. Teeple, A. S. Navathe, *Addressing Bias in Artificial Intelligence*, in *Health Care*, JAMA, 2019.
[130] In merito di segnala il contributo offerto dal Garante per la protezione dei dati personali che, in data 10.10.2023, ha diramato il *Decalogo per la realizzazione di servizi sanitari nazionali attraverso sistemi di Intelligenza Artificiale*, doc. web n. 9938038.
[131] O. Colpani, *Machine Learning…*, *op. cit.*, p. 5-11, che rinvia a A. Rajkomar et al., *Machine Learning in Medicine*, in N. *Engl. J. Med.*, 380, 2019, pp. 1347-1358.
[132] N. Musacchio, F. Baccetti, C.B. Giorda, G. Guaita, L. Morviducci, B. Nreu, A. Ozzello, P. Ponzani, A. Rossi, R. Zilich, *Intelligenza artificiale: la posizione e le strategie di AMD – Artificial Intelligence: AMD position and strategies*, in *The Journal of AMD, Associazione Medici Diabetologi*, vol. 25, n. 4, 2022, p. 219 ss.
[133] M. Muselli, *Extracting knowledge from biomedical data through Logic Learning Machines and Rulex*, in *Technological-industrial Communications*, 2012.
[134] N. Musacchio, F. Baccetti, C.B. Giorda, G. Guaita, L. Morviducci, B. Nreu, A. Ozzello, P. Ponzani, A. Rossi, R. Zilich, *Intelligenza artificiale…*, *op. cit.*, p. 219 ss.

4. 'What we could do differently' (prescriptive analytics). Ovvero: quali sono gli aspetti su cui si potrebbe
intervenire per evitare che succeda una situazione che è meglio evitare?
5. 'What-if' (analisi d'impatto – simulazione di scenari). Ovvero: se nel comportamento passato ci si fosse comportati diversamente, quale impatto, o beneficio, si sarebbe potuto generare sugli esiti attesi? E con quale probabilità?

Siffatte applicazioni di IA comportano, ovviamente, il trattamento dei dati "sanitari" che vengono raccolti nel tempo attraverso il Fascicolo sanitario elettronico (FSE), il Dossier sanitario e la cartella clinica elettronica, per essere aggregati e dematerializzati nei *repository* di ASL e Aziende Ospedaliere[135]. Di conseguenza, è opportuno fare brevi cenni ai profili riguardanti la *privacy* di tali dati. In tal senso, va ricordato che l'art. 9, comma 1, lettera b), del D.L. 8 ottobre 2021, n. 139, convertito, con modificazioni, dalla L. 3 dicembre 2021, n. 205, ha introdotto il comma 1-*bis* all'art. 2-*sexies* del Codice della privacy. Esso ha previsto che i dati personali relativi alla salute, privati di elementi identificativi diretti, possano essere trattati dal Ministero della salute e da altri enti tra cui le regioni/province autonome, relativamente ai propri assistiti, anche mediante l'interconnessione a livello nazionale dei sistemi informativi su base individuale del Servizio sanitario nazionale, ivi incluso il Fascicolo sanitario elettronico, aventi finalità compatibili con quelle sottese al trattamento, con le modalità e per le finalità che saranno stabilite con decreto del Ministro della salute, che dovrà essere adottato previo parere del Garante e nel rispetto di quanto previsto dal Regolamento, dal Codice, dal Codice dell'amministrazione digitale e dalle linee guida dell'Agenzia per l'Italia digitale, in materia di interoperabilità.

5. Aspetti normativi in materia di tutela dei dati personali utilizzati nell'IA: l'explainability dell'algoritmo.

[135] Nel settore sanitario, la raccolta di *big data* aggregati consente di creare mappe individuali e collettive, riferite sia al singolo paziente che al sistema sanitario nel suo complesso, tuttavia, trattandosi di una «ingente mole di informazioni, provenienti da plurime fonti e idonee a perseguire molteplici finalità, per il tramite di tecniche informatizzate (c.d. software o algoritmi) […] si riferiscono a grandi set di dati raccolti, periodicamente o automaticamente, che vengono archiviati elettronicamente, riutilizzabili allo scopo di migliorare le prestazioni del sistema sanitario» (Commissione UE, Study on Big Data in Public Health, Telemedine and Healthcare. Final Report, december 2016).

In generale, la tutela dei dati personali nei trattamenti svolti mediante strumenti algoritmici rappresenta uno dei tanti principi di *governance* dell'IA. Nelle già accennate Linee Guida *Ethics guidelines for trustworthy AI* del 2019 (v. par. 2 del presente contributo), la Commissione UE affermava che «La trasparenza è un requisito fondamentale per garantire che il sistema di AI non sia viziato da *bias*. I sistemi di Intelligenza Artificiale utilizzati per prendere decisioni che riguardano gli individui sono soggetti al principio di 'diritto alla spiegazione': devono poter esplicitare la logica di funzionamento dei modelli che hanno generato»[136]. Il requisito della trasparenza dei sistemi di IA è strettamente connesso al principio della spiegabilità e comprende la trasparenza degli elementi pertinenti: i dati, il sistema e i modelli di business[137]. La spiegabilità è possibile se sono sufficientemente "tracciabili" e "documentabili" i *set* di dati e i processi che determinano la decisione del sistema di IA, compresi quelli di raccolta ed etichettatura dei dati, come pure gli algoritmi utilizzati[138]. In particolare, è "spiegabile" un sistema di IA se sono spiegabili sia i suoi processi tecnici sia le relative decisioni umane[139]. Infine, la Commissione auspica che i sistemi di IA siano identificabili come tali dagli esseri umani, i quali devono sempre essere a conoscenza del fatto che stanno interagendo con un sistema di IA[140]. Nell'IA, il tema della responsabilità è connesso con i temi della sicurezza, della trasparenza e della spiegabilità, del controllo umano. In dottrina si è affermato che la responsabilità è uno strumento per avere un'IA spiegabile, capace di conquistare la fiducia del pubblico e dissiparne i timori[141]. Si avverte l'esigenza, secondo il *World Health Organization*, di un compromesso tra spiegabilità e accuratezza degli algoritmi. Ad esempio, in ambito medico, ad un paziente si può spiegare l'utilità di una terapia rispetto a un'altra, mentre non è sempre possibile spiegare come funziona il meccanismo di azione di quella terapia. Allo stesso modo, è più importante spiegare come un sistema di IA sia stato validato e l'affidabilità dei suoi risultati, piuttosto che spiegare come un modello di IA arriva

[136] High-Level Expert Group on AI, Gruppo di esperti ad alto livello sull'intelligenza artificiale, *The Ethics Guidelines for Trustworthy Artificial Intelligence (AI), Orientamenti etici per un'IA affidabile*, 2019, https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[137] *Ibidem*, 75.

[138] *Ibidem*, 76.

[139] *Ibidem*, 77.

[140] *Ibidem*, 78.

[141] J. Fjeld, N. Achten, H. Hilligoss, A. C. Nagy, M. Srikumar, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, in *The Berkman Klein Center for Internet & Society Research Publication Series*, 1, 2020, https://cyber.harvard.edu/publication/2020/principled-ai

a una particolare decisione[142]. I casi in cui sono in pericolo i diritti umani devono essere sempre identificabili. In altre parole, deve essere trasparente in quale contesto il sistema algoritmico è stato implementato e per quale scopo. La spiegabilità delle scelte algoritmiche è emersa con l'emanazione del GDPR, e, soprattutto, con la pubblicazione delle considerazioni di Goodman e Flaxman, in cui gli stessi hanno affermato che il GDPR avrebbe introdotto un diritto a una "spiegazione" nel combinato disposto tra gli artt. 13 e 14, dedicati agli obblighi informativi del titolare del trattamento[143], e l'art. 22 («Diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che riguardano l'interessato che incida in modo analogo significativamente sulla sua persona»), e, di conseguenza, gli interessati dovrebbero poter ottenere l'*explainability* di una decisione algoritmica che è stata presa su di essi, e, più in particolare, di come i sistemi di IA abbiano raggiunto una determina decisione, al fine di metterli in condizione di valutare, e, potenzialmente, correggere, tali scelte[144]. Contrariamente, altri osservatori sono dell'avviso che il GDPR avrebbe introdotto soltanto degli "obblighi di notifica", per i produttori e/o fornitori di sistemi di IA, nel momento in cui i dati vengano raccolti[145]. Appare evidente come il principio di trasparenza e spiegabilità dell'algoritmo coinvolga anche il principio di tutela della privacy nell'IA e la questione non è di poco conto, poiché, se il titolare rileva da terzi l'algoritmo, non potrà darne conto, con conseguente violazione del rispetto dei principi di requisiti di liceità, di correttezza e di trasparenza del trattamento, sanciti nel GDPR[146]. È altresì importante il contestato par. 3 dell'art. 22 del GDPR, laddove specifica l'obbligo del titolare di garantire all'interessato di

[142] World Health Organization, *Ethics and governance of artificial intelligence for health: WHO guidance*, 2021.

[143] L'articolo 22 del GDPR va letto di concerto con altre due norme, ossia l'art. 13, par. 2, lettera f) e l'art. 14, par. 2, lettera g), GDPR, laddove inseriscono, tra le informazioni che il titolare del trattamento è obbligato a fornire all'interessato, quella sull'eventuale «esistenza di un processo decisionale automatizzato, compresa la profilazione di cui all'articolo 22, paragrafi 1 e 4, e, almeno in tali casi, informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato». Per informative significative si intende la comunicazione delle finalità, le modalità mediante le quali avviene la profilazione, la logica pensata e le conseguenze previste per l'interessato a seguito di tale tipo di trattamento, intendendo, in tal senso, i criteri utilizzati per giungere alla decisione, senza, tuttavia, dover necessariamente fornire una spiegazione complessa degli algoritmi utilizzati o dover garantire l'ostensione dell'algoritmo completo. In proposito, cfr. anche R. Mortier, *Human data interaction: The human face of the data-driven society*, in *MIT Technology Review*, 2014.

[144] B. Goodman, S. Flaxman, *European Union Regulations on algoritmic decision making and a "right to Explanation*, in *Association for the Advancement of Artificial Intelligence*, 2016.

[145] S. Wachter, B. Mittelstadt, L. Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International data privacy laws*, 7, 2, 2017.

[146] F. Lorè, *Una Intelligenza Artificiale più umana, tra etica e privacy*, in *Ciberspazio e Diritto*, 22, 67, 1, 2021.

«esprimere la propria opinione e di contestare la decisione» automatizzata, poiché coordinata tale disposizione con le future regole del Reg. UE sull'IA, appare sproporzionata per alcuni settori applicativi dell'IA, e insufficiente per altri, in particolare per il settore delle reti neurali e, alla luce del Considerando 51 del GDPR[147], e per il settore del riconoscimento facciale, attesa la prescrizione di non poter basare, le decisioni di cui all'art. 22, su categorie "particolari" di dati personali, tra cui rientrano quelli biometrici. Inoltre, l'art. 15, par. 1, del GDPR, garantisce all'interessato il diritto di ottenere dal titolare del trattamento la conferma che sia o meno in corso un trattamento di dati personali che lo riguardano e in tal caso, di ottenere l'accesso ai dati personali. Alla lettera h), in particolare, l'art. 15 concede la possibilità di conoscere «l'esistenza di un processo decisionale automatizzato, compresa la profilazione di cui all'articolo 22, paragrafi 1 e 4, e, almeno in tali casi, informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato». Su siffatto diritto a una spiegazione dell'algoritmo, come risultante dal combinato disposto delle norme cui si è fatto cenno, non vi è unanimità di posizioni tra gli osservatori, tuttavia, anche la dottrina più rigida ha ritenuto esistente il diritto di ottenere informazioni generali sulla funzionalità dell'algoritmo e sulle conseguenze previste da tale operazione sui dati. Secondo Wachter, Mittelsttad e Floridi, infatti, alla luce dell'art. 22, l'obbligo della spiegazione dei meccanismi decisionali dell'algoritmo opererebbe non *ex ante* rispetto al trattamento (non essendo ancora stata presa alcuna decisione sulla base di trattamenti automatizzati), ma *ex post*, ovvero soltanto in caso di effettivo esercizio del diritto di accesso da parte dell'interessato[148]. Wachter, Mittelsttad e Floridi sostengono che gli artt. 13 e 14 del Regolamento creerebbero soltanto degli obblighi "di notifica" per i titolari e i responsabili del trattamento nel momento in cui i dati vengano raccolti, con il conseguente obbligo, per il titolare, di informare l'utente dell'esistenza di un processo decisionale automatizzato e di fornire spiegazioni significative sulla logica sottostante, sul significato e sulle conseguenze previste del trattamento. Le argomentazioni di quest'ultima dottrina appaiono, a chi scrive,

[147] «Il trattamento di fotografie non dovrebbe costituire sistematicamente un trattamento di categorie particolari di dati personali, poiché esse rientrano nella definizione di dati biometrici soltanto quando saranno trattate attraverso un dispositivo tecnico specifico che consente l'identificazione univoca o l'autenticazione di una persona fisica».

[148] S. Wachter, B. Mittelstadt, L. Floridi, *Why a Right…*, *op. cit.* Cfr. anche E. Troisi, *AI e GDPR: l'Automated Decision Making, la protezione dei dati e il diritto alla 'intellegibilità dell'algoritmo*, in *European Journal of Privacy Law & Technologies*, 1, 2019, http://www.ejplt.tatodpr.eu/Article/Archive/index_html?ida=64&idn=4&idi=-1&idu=-1

ragionevoli e concretamente attuabili, e sembrano essere sussumibili nelle affermazioni dei garanti europei, secondo cui gli articoli 13 e 14 del GDPR hanno un ambito di applicazione più ampio di quello di cui all'art. 22, di conseguenza l'interessato ha in ogni caso diritto ad una spiegazione dell'esistenza di trattamento automatizzato di dati (seppur con minor dettaglio quando non si tratti di decisioni basate «unicamente sul trattamento automatizzato»)[149]. Per completezza, è opportuno segnalare altresì la recente guida pubblicata dall'autorità di controllo spagnola in merito al principio di trasparenza del trattamento dei dati di cui agli articoli 12 e seguenti del *GDPR*, che, come accennato, in alcuni casi si sovrappone al principio di trasparenza algoritmica come enucleato nell'*AI Act*. Ove i sistemi di IA utilizzino dati personali (nella progettazione o nello sviluppo) o conservino ed elaborino dati di persone identificate o identificabili, devono essere conformi al GDPR, in particolare attraverso una DPIA, una informativa privacy e una procedura di gestione dell'esercizio dei diritti da parte degli interessati. D'altro canto, l'autorità spagnola ha ricordato che i sistemi di IA devono essere progettati in modo da consentire la trasparenza e l'audit, al fine di evitare *bias* e discriminazioni. L'Autorità ha difatti affermato che il termine "trasparenza" nell'*AI Act* viene utilizzato con un significato diverso rispetto a quello del *GDPR*, poiché gli obblighi in materia di trasparenza del sistema di IA sono applicabili fin dalla sua progettazione e durante tutto il suo ciclo di vita, indipendentemente dal fatto che vengano trattati dei dati personali[150]. La trasparenza dell'*AI Act* è legata alla spiegabilità, alla documentazione, alla tenuta dei registri e alla fornitura di informazioni sulle modalità di utilizzo di detto sistema di IA, al fine di rendere edotti gli utenti su tali circostanze. Con riferimento, in particolare, all'ambito della medicina, va altresì ricordato che il Consiglio d'Europa, allo scopo di evitare pericolosi automatismi, ha ribadito la necessità di un approccio *human in command* e l'irrinunciabilità dell'apporto umano nell'erogazione delle cure mediche[151]. Eppure, il limite, sotto il profilo giuridico, è costituito proprio da tale

[149] WP29, Gruppo di lavoro Art. 29, *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01), Linee guida in materia di processi automatizzati e profilazione*, 3 ottobre 2017, nella versione emendata il 6 febbraio 2018, https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

[150] Agencia Española de Protección de Datos, *Inteligencia artificial: Transparencia*, 20 settembre 2023, https://www.aepd.es/prensa-y-comunicacion/blog/inteligencia-artificial-transparencia

[151] F. Ferretti, *Intelligenza artificiale…*, op. cit. Cfr. anche B. Mittelstadt, *The impact of Artificial Intelligence on the doctor-patient relationship*, Report commissionato dallo Steering Committee for Human Rights in the field of Biomedicine and Health (CDBIO), Consiglio d'Europa, dicembre 2021, p. 52: «The Council of Europe has clearly recognised the risk of automation bias in calling for guarantees that AI-driven health applications do not replace

aspetto. Le informazioni sul trattamento dei dati personali da fornire al paziente riguardano, infatti, anche gli strumenti di IA eventualmente coinvolti, tuttavia, le disposizioni degli articoli 13, 14, e 22 del GDPR, sono applicabili soltanto se la decisione sia totalmente delegata all'algoritmo, non essendo prevista una "spiegazione" delle decisioni prese dal medico. Nella Proposta di *AI Act* europeo è rinvenibile una norma che obbliga i fornitori a garantire che i sistemi di IA destinati a interagire con le persone fisiche «siano progettati e sviluppati in modo tale che le persone fisiche siano informate del fatto di stare interagendo con un sistema di IA, a meno che ciò non risulti evidente dal punto di vista di una persona fisica ragionevolmente informata, attenta e avveduta, tenendo conto delle circostanze e del contesto di utilizzo»[152]. Tali prescrizioni, tuttavia, hanno carattere generale, e difficilmente potranno essere applicate ad algoritmi predittivi e decisionali. E non va dimenticato che le criticità accennate costituiscono soltanto uno dei numerosi temi da affrontare, come, ad esempio, la sicurezza del dispositivo medico, ai sensi dell'art. 32, GDPR, o l'impiego, nel *training* dei sistemi algoritmici, di *set* di dati che rientrano nella nozione di "categorie particolari di dati personali", di cui all'art. 9, GDPR[153]. Per ragioni di spazio, non è possibile, in questa sede, fare delle considerazioni sull'uso di *big data* (o *big datasets*) nell'IA, basterà ricordare che ad essi non si dovrebbe applicare il GDPR poiché sarebbero *ab origine* anonimi[154], sebbene sia necessario verificare se realmente l'interessato non sia più identificabile in alcun modo, poiché spesso l'anonimizzazione viene confusa con la pseudonimizzazione, operazione, quest'ultima, che non sempre pone il trattamento dei dati al di fuori del perimetro di applicazione del GDPR[155]. Non a caso, anche l'Autorità garante per la protezione dei

---

human judgement completely and that thus enabled decisions in professional health care are always validated by adequately trained health professionals».

[152] Art. 52, paragrafi 1 e 2, della Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale, *AI Act*) e modifica alcuni atti legislativi dell'Unione, 21 aprile 2021 (2021/0106 (COD)), di cui è prevista l'adozione definitiva, all'esito dei "triloghi" tra le istituzioni europee, entro la fine del 2023.

[153] L. Scaffardi, *La medicina alla prova dell'Intelligenza Artificiale*, in *DPCE online*, 1, 2022, p. 349 ss.

[154] E. Errigo, *Big healthcare data sets: circolazione e nuove forme di appartenenza*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2022; G.M. Ruotolo, *I dati non personali: l'emersione dei big data nel diritto dell'Unione europea*, in *Studi sull'integrazione europea*, XIII, 2018, p. 97 ss.; K. Karaboue, *Intelligenza artificiale nell'ambito del sistema sanitario. Implicazioni in termini di privacy alla luce del nuovo GDPR*, in *Amministrativ@mente*, 1, 2023.

[155] P. Musacchio, *L'adeguamento dei siti web alle Linee Guida sui cookie (2021) del Garante privacy*, in *Ratio Juris*, LXXI, novembre 2021, https://www.ratioiuris.it/ladeguamento-dei-siti-web-alle-linee-guida-sui-cookie-2021-del-garante-privacy/; WP29, Gruppo di lavoro Art. 29, Parere 05/2014 sulle tecniche di anonimizzazione, p. 7. De-identificare significa eliminare la correlazione tra i dati personali e una determinata persona fisica, rendendo impossibile la sua identificazione. Se l'anonimizzazione va a buon fine, i dati oggetto dell'operazione non sono

dati personali ha iniziato a occuparsi del tema dell'IA in medicina, costituendo, peraltro, un gruppo di Lavoro Interistituzionale sulla medicina predittiva[156] per la predisposizione del Regolamento attuativo dell'articolo 8 del decreto legge 19 maggio 2020, n. 34, convertito con modificazioni in legge 17 luglio 2020, n. 77, tenuto a individuare «i dati personali…che possono essere trattati, le operazioni eseguibili, le modalità di acquisizione dei dati dai sistemi informativi che li detengono e le strutture appropriate e specifiche per tutelare i diritti degli interessati, nonché i tempi di conservazione dei dati trattati». In una pronuncia del 2020, l'Autorità di controllo nazionale, nel fornire un parere al Consiglio di Stato sulle nuove modalità di ripartizione del fondo sanitario tra le regioni proposte dal Ministero della salute e basate sulla stratificazione della popolazione, ha precisato, innanzitutto, che la profilazione del rischio sanitario individuale e la successiva stratificazione della popolazione per classi di morbilità attraverso l'uso di algoritmi richiede come base giuridica soltanto un provvedimento normativo[157]. Tale intervento del legislatore si è realizzato nell'art. 7, comma 1, del richiamato decreto legge 19 maggio 2020, n. 34, laddove ha previsto che il Ministero della salute, nell'ambito dei compiti e delle funzioni istituzionali in materia di prevenzione, diagnosi, cura e riabilitazione delle malattie, nonché di programmazione tecnico sanitaria di rilievo nazionale e indirizzo, coordinamento, monitoraggio dell'attività tecnico sanitaria regionale, può trattare, ai sensi dell'articolo 2-*sexies*, comma 2, lettera v) del Codice privacy, i dati personali anche relativi alla salute, degli assistiti, raccolti nei sistemi informativi del Servizio sanitario nazionale, per lo sviluppo di metodologie predittive dell'evoluzione del fabbisogno di salute della popolazione. Il comma secondo della medesima disposizione stabilisce che con decreto del Ministero, di natura non regolamentare, siano individuati i dati personali, anche inerenti alle categorie particolari, che possono essere trattati a tal fine, le operazioni eseguibili, le modalità di acquisizione dei dati dai sistemi informativi dei soggetti che li detengono e le misure appropriate e specifiche per tutelare i diritti degli interessati, nonché i tempi di conservazione dei dati trattati. Di concerto, il Garante ha posto come riferimenti fondamentali, per l'uso dell'IA nei procedimenti amministrativi, i tre principi affermati dal Consiglio di Stato nel 2019, ossia quello della "conoscibilità", per cui

più classificati come dati personali, e quindi non rientrano più nell'applicazione del GDPR, fermo restando che fino a tale momento, l'anonimizzazione è pur sempre un'operazione di trattamento di dati personali.

[156] cui sono state invitate a partecipare anche otto regioni/province autonome (Lazio, Emilia-Romagna, Lombardia, Piemonte, Puglia, Toscana e Veneto e alla Provincia Autonoma di Bolzano).

[157] Garante privacy, Provvedimento 5 marzo 2020, n. 43.

ognuno ha diritto a conoscere l'esistenza di processi decisionali automatizzati che lo riguardino ed in questo caso a ricevere informazioni significative sulla logica utilizzata, quello della "non esclusività" della decisione algoritmica, alla luce dell'art. 22, GDPR, nonché quello della "non discriminazione" algoritmica, per far sì che sia garantita l'assenza di effetti discriminatori nei confronti delle persone sulla base della razza o dell'origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell'appartenenza sindacale, dello status genetico, dello stato di salute o dell'orientamento sessuale[158]. L'applicazione di tali principi all'ambito della medicina implica la necessità di progettare un'IA capace di considerare con maggiore profondità le differenze all'interno di ciascun *cluster* di assistiti, con la conseguenza dell'emersione dell'esigenza di svolgere degli *audit* dei sistemi algoritmici (verificabilità), considerato che sussiste il rischio che l'IA potrebbe condurre a distorsioni dell'azione amministrativa, tema che per ragioni di spazio non può essere qui trattato[159]. Inoltre, con riferimento ai profili etici, rilevanti "lato normativa privacy", è opportuno segnalare, e lo ha fatto anche l'Autorità garante nel 2020, che nell'implementazione di sistemi di IA destinati all'ambito sanitario, è necessario procedere, preventivamente, a un *Data Protection Impact Assessment* (*DPIA*) del trattamento, come vuole, d'altronde, l'art. 35 del GDPR[160]. Il Garante ha anche affermato, nello stesso Provvedimento, che l'attività di profilazione in campo medico non può prescindere dal consenso dell'interessato, essendo un trattamento autonomo e ulteriore rispetto alle finalità di cura[161]. Infine, va rilevato un recente ammonimento del Garante privacy comminato alla Regione Lazio per illiceità del trattamento dei dati personali effettuato dal medesimo Ente, per la violazione del principio di liceità, correttezza e trasparenza del trattamento, di cui all'art. 5, par. 1, lett. a) del Regolamento e degli artt. 6, 9 e 35 del Regolamento[162]. Il Provvedimento consegue

[158] Consiglio di Stato, Sez. VI, 13 dicembre 2019, n. 8472.

[159] In merito, cfr., *ex amplius*, F. Conte, *La trasformazione digitale della pubblica amministrazione: il processo di transizione verso l'amministrazione algoritmica*, in *federalismi.it*, 11, 2023; L. Musselli, *La decisione amministrativa nell'età degli algoritmi: primi spunti*, in *MEDIALAWS - Rivista di Diritto dei Media*, n. 1, 2020; P. Forte, *Diritto amministrativo e data science. appunti di intelligenza amministrativa artificiale (AAI)*, in *PA Persona e amministrazione, Ricerche Giuridiche sull'Amministrazione e l'Economia*, 1, 2020; E. Carloni, *Algoritmi su carta: politiche di digitalizzazione e trasformazione digitale delle amministrazioni*, in *Diritto pubblico*, 2019; D.-U. Galetta, J. G. Corvalán, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *federalismi.it*, 3, 2019; S. Vaccari, *Note minime in tema di Intelligenza Artificiale e decisioni amministrative*, in *Giustamm*, 10, 2019; L. Viola, *L'intelligenza artificiale nel procedimento e nel processo amministrativo: lo stato dell'arte*, in *federalismi.it*, 21, 2018.

[160] Garante privacy, Provvedimento 8 maggio 2020, n. 84.

[161] Garante privacy, *Ibidem*.

[162] Garante privacy, Provvedimento 24 febbraio 2022, n. 64.

all'attività svolta dal richiamato gruppo interistituzionale (si veda il paragrafo 4 della presente trattazione), volta all'elaborazione della versione prototipale del modello predittivo, idoneo a valutare il fabbisogno della popolazione italiana, attraverso la profilazione del rischio sanitario individuale e la successiva stratificazione per classi di morbilità della stessa, cui sono state invitate a partecipare anche otto regioni/province autonome. Di conseguenza, l'Ufficio competente dell'Autorità ha richiesto informazioni al Ministero della salute, al fine di valutare la conformità alla disciplina vigente in materia di protezione dei dati personali delle attività in corso per la realizzazione della suddetta versione prototipale del modello predittivo. Poiché il Ministero ha rappresentato di aver "recepito esclusivamente elaborati statistici già aggregati da (…) otto Regioni italiane" (tra cui la Regione Lazio), che "hanno agito quali autonomi titolari del trattamento, nell'ambito di proprie attività istituzionali, che prevedono l'interconnessione e la successiva aggregazione per scopi di ricerca scientifica, nonché di programmazione, gestione, controllo e valutazione dell'assistenza sanitaria", il Garante ha interpellato le Regioni. Orbene, la Regione Lazio ha precisato che "non sussistono dati identificativi diretti e che i dati sono pseudonimizzati alla fonte a mezzo dei noti meccanismi di cifratura applicati dalle Regioni" e che "il Ministero ha identificato criteri di aggregazione aderenti ad una soglia minima di frequenza per ogni combinazione pari a 3, anche in accordo all'art. 4, Allegato 1, delle "Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica pubblicate ai sensi dell'art. 20, comma 4, del d.lgs. 10 agosto 2018, n. 101 - 19 dicembre 2018"; ciò al fine di garantire l'impossibilità di identificazione, anche indiretta, delle persone fisiche con l'aggregazione dei dati sopra definita". Ancora, l'ente laziale ha affermato che non vi è stata possibilità alcuna per il Ministero di reidentificare gli interessati, giusta la natura pseudonimizzata dai dati relativi alla salute trattati dalla Regione e la natura aggregata e anonima delle informazioni successivamente rese al Ministero. Il Garante, tuttavia, è di diverso avviso. Anzitutto, secondo l'autorità di controllo, la Regione «ha effettuato l'elaborazione e la successiva aggregazione di dati presenti nei sistemi informativi sanitari che detiene in qualità di titolare trattamento ex lege, in assenza di un idoneo presupposto giuridico (art. 4, n. 7 del Regolamento e Guidelines 07/2020 on the concepts of controller and processor in the GDPR, punto 21). La circostanza che un soggetto terzo, nel caso in esame rappresentato dal Ministero della salute (cfr. art. 4, n. 10 del Regolamento), chieda a un titolare (Regione Lazio) di effettuare operazioni di trattamento su dati personali rispetto ai quali quest'ultimo è titolare, indicandone

anche le modalità, non comporta infatti l'automatica attribuzione della titolarità in capo al richiedente, né tantomeno la perdita della titolarità da parte del soggetto che legittimamente detiene i dati. Spetta
a quest'ultimo invero valutare la legittimità della richiesta e in particolare la sussistenza di una idonea base giuridica per effettuare le operazioni di trattamento richieste, tanto più che, nel caso di specie, le predette operazioni hanno riguardato dati sulla salute di tutta la popolazione assistita regionale attraverso l'uso di algoritmi». Peraltro, il garante ricorda che quando sono stati effettuati i trattamenti da parte della Regione, non risultava ancora vigente il comma 2-bis dell'art. 7 del d.l. 34 2020, introdotto con il richiamato d.l. 139 del 2021. Sulla base dell'erroneo presupposto di non essere titolare del trattamento, la Regione non ha svolto alcuna valutazione di impatto ai sensi dell'art. 35 del Regolamento, ritenendo che tale compito spettasse al Ministero. La circostanza che il trattamento *de quo* sia stato richiesto dal Ministero non può costituire, secondo il Garante, una condizione esimente dall'obbligo di effettuare una valutazione di impatto, posto che la scelta di dar seguito alla richiesta del Ministero è stata operata dalla Regione, che, a tal fine, ha elaborato i dati personali di cui è titolare.

6. Conclusioni

Al netto delle elucubrazioni sulla responsabilità civilistica delle scelte dei sistemi algoritmici, tema ampiamente affrontato in questa trattazione, in dottrina si è correttamente osservato come una errata configurazione dell'IA possa porre in serio rischio fondamentali interessi individuali e collettivi[163], a causa, come accennato, dei *bias* algoritmici, in grado di mettere in pericolo i principi di eguaglianza e non discriminazione[164]. L'uso dell'IA, infatti, implica una serie di scelte poco oggettive e neutre, l'adozione di schemi predittivi e di criteri capaci di portare alla formulazione di giudizi, frutto di precise scelte e di valori, in maniera consapevole o inconsapevole[165]. I casi affrontati dalla giurisprudenza sono ancora pochi, a partire dai più famosi, oltreoceano, *Loomis v. Wisconsin*[166] e *Ewert v. Canada*[167], e, in Italia,

[163] G. Sartor, F. Lagioia, *Le decisioni algoritmiche tra etica e diritto*, in: U. Ruffolo (a cura di), *Intelligenza artificiale. Il diritto, i diritti…*, *op. cit.*, p. 63 ss.

[164] A. Simoncini, *L'algoritmo incostituzionale…*, *op. cit.*

[165] N. Muciaccia, *Algoritmi e procedimento decisionale: alcuni recenti arresti della giustizia amministrativa*, in: *federalismi.it*, 2, 2020, che riprende P. Schwartz, *Data Processing and Government Administration: The Failure of the American Legal Response to the Computer*, in *43 Hastings L.J.*, 1992, p. 1341 ss.

[166] *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), con nota in *Harvard Law Review*, 130, 2017, p. 1530 ss.

quello relativo all'algoritmo di assegnazione dei turni di lavoro dei *rider* della piattaforma *Deliveroo*[168]. Come accennato, per rispondere agli interrogativi epistemologici sulla natura della conoscenza e sulla riproducibilità dell'intelligenza umana si è auspicata la centralità del rispetto dei principi della *governance* dell'IA nell'operare scelte eticamente orientate[169]. Uno di tali principi è costituito dalla responsabilità (*accountability)* dei sistemi di IA – intesa sia come *liability*, sia come *accountability* – che non può risiedere nella tecnologia stessa, ma deve essere ripartita tra coloro che la progettano, la sviluppano e la implementano (Responsabilità di per sé, *Accountability Per Se*)[170]. Connesso al principio di *accountability* dell'IA, così come al principio di accuratezza (*accuracy*) e qualità (*quality*) dei dataset degli algoritmi, si colloca quello di equità (*equity*), considerato che una delle criticità più rilevanti è data dalla rappresentatività dei dati di *input* (quello che viene definito in gergo "*garbage in, garbage out*"). Tali considerazioni valgono, tanto più, per l'IA "generativa" - ramo che non sembra emergere in alcuna definizione specifica, al livello normativo – le cui applicazioni sono accomunate da un elemento caratteristico quale il suo meccanismo di autocompletamento, poiché riesce, come spiegato in un fulgido esempio dalla dottrina recente, «in forza dei dati che sono stati inseriti (ad es. l'*input* testuale di una storia o la richiesta di esecuzione di un comando del tipo "Scrivi un programma che calcoli la media di questa distribuzione di dati"), a completarli in modo plausibile (e il più delle volte corretto)» in una sorta di «partenogenesi o di gemmazione»[171]. Ai rischi analizzati nella presente trattazione si aggiunge quello di non riuscire più, in un lontano futuro, a riconoscere se si sta interagendo con un'IA e a comprendere il

---

[167] *Ewert v. Canada*, 2018 SCC 30 [2018] 2 S.C.R. 165.

[168] Tribunale di Bologna, Sezione Lavoro, ordinanza 31 dicembre 2020, in cui il Giudice del Lavoro ha riconosciuto la natura discriminatoria delle strategie di profilazione dei rider adottate, con conseguente condanna a risarcire le organizzazioni sindacali ricorrenti e a dare ampia pubblicità al provvedimento.

[169] D.-U. Galetta, *Human-stupidity-in-the-loop?..., op. cit.*

[170] J. Fjeld, N. Achten, H. Hilligoss, A. C. Nagy, M. Srikumar, *Principled Artificial Intelligence…, op. cit.*, studio in cui è stato analizzato un gran numero di documenti sull'IA, appurando che quasi tutti menzionano almeno un principio di accountability: "adozione di nuove normative", "verificabilità e replicabilità", "valutazioni di impatto", "responsabilità ambientale", "requisiti di valutazione e audit", "creazione di un organismo di monitoraggio", "capacità di presentare ricorso", "rimedio per decisione automatizzata", "responsabilità e responsabilità legale", "responsabilità di per sé". Tali principi sono stati mappati attraverso il ciclo di vita di un sistema di IA, in tre fasi essenziali: progettazione (pre-distribuzione o design), che racchiude verificabilità e replicabilità, valutazione di impatto e responsabilità ambientale; monitoraggio (durante la distribuzione), che comprende requisiti di valutazione e audit, creazione di un organismo di vigilanza, capacità di ricorso; riparazione (dopo che si è verificato un danno), che comprende rimedi per decisioni automatizzate, responsabilità legale, adozione di nuove regolamentazioni.

[171] I. Coppola, *Intelligenza artificiale generativa: GPT o chatGPT verso ipotesi di ragionamento automatici e linguaggio artificiale. Tra l'argomentazione giuridica di Bobbio ed il principio di precauzione*, in *Diritto di internet*, 30 marzo 2023.

suo funzionamento, finendo per non essere più in grado di sovvertirne le scelte. L'appiattimento culturale che ne deriverebbe "farebbe gola" alle autocrazie e sarebbe in grado di condizionare pesantemente l'opinione pubblica, senza dimenticare il potenziale aumento delle diseguaglianze e dei licenziamenti che potrebbe produrre. In tale contesto, considerato che l'evoluzione tecnologica dei modelli di IA porterà a cambiamenti dirompenti nel modo di sviluppare applicazioni, nella concentrazione di potere tecnologico e nella disparità tra i detentori della tecnologia e gli altri, e infine nel mondo delle professioni - come osservato correttamente in dottrina – probabilmente, oggi, l'errore più grande sarebbe quello di lasciare l'uso (e dunque il controllo) dell'IA a poche grandi imprese che detengono le enormi risorse necessarie per costruire i modelli più sofisticati[172]. Si pensi soltanto, a titolo esemplificativo, che *Microsoft* ha annunciato di volere investire 10 miliardi di dollari in *OpenAI* e nell'integrazione di *ChatGPT* con il suo motore di ricerca *Bing*, e che, nel marzo scorso, al lancio di *Bard* (la *ChatGPT* di *Google*) una notizia inesatta di astronomia è costata, in borsa, 100 miliardi di dollari (il 7%). Viene facile chiedersi, dunque, cosa sarebbe successo se l'errore si fosse verificato nel campo della politica, della medicina o, ancora, del cambiamento climatico. Appare chiaro come le *chatbot* possano divenire un micidiale strumento di disinformazione. Sotto il profilo della disciplina, oltre alle disposizioni del GDPR in merito alla tutela dei dati personali, che vanno ad alimentare il *training* dei sistemi algoritmici, e al netto delle accennate contromisure previste dal Regolamento UE sull'IA di prossima adozione, emerge anche la soluzione normativa individuata dal legislatore europeo nel *Digital Service Act*[173], entrato in vigore nel 2022, ovvero quella, in sostanza, di una sorta di *auditing* degli algoritmi in merito a tre situazioni giuridiche emergenti, ossia il diritto di conoscere in ogni circostanza la natura artificiale dei sistemi di IA, il diritto a una spiegazione

[172] G. Attardi, *I limiti e i rischi di ChatGPT non sono quelli che crediamo, possiamo ancora evitare il peggio*, in *I quaderni di Agenda Digitale*, 13, 2023, p. 33 ss.

[173] Nell'art. 12 del *DSA* è stabilito che i prestatori di servizi intermediari includono nelle loro condizioni generali informazioni sulle restrizioni che impongono in relazione all'uso dei loro servizi per quanto riguarda le informazioni fornite dai destinatari del servizio. Tali informazioni, che riguardano, tra l'altro, il processo decisionale algoritmico e la verifica umana, devono essere redatte, nelle condizioni generali, in un linguaggio chiaro e privo di ambiguità, e disponibili al pubblico in un formato facilmente accessibile. Importantissima è, infine, la puntualizzazione del paragrafo 2, laddove impone ai prestatori di servizi intermediari di agire in modo diligente, obiettivo e proporzionato nell'applicare e far rispettare le restrizioni, tenendo conto dei diritti fondamentali applicabili del servizio. Tali obblighi, dunque, sono cogenti per le piattaforme, anche private, che utilizzano algoritmi.

degli *output* di tali sistemi, nonché il diritto a un controllo umano su di essi[174]. Va evidenziato altresì che il *DSA* individua il fulcro delle problematiche e dei rischi legati alla mancanza di trasparenza nelle operazioni algoritmiche di classificazione e nella selezione dei contenuti. Sono pertanto oggetto di attenzione quei sistemi di IA che «sfruttano notifiche abusive o altri metodi per ostacolare la concorrenza o mettere a tacere l'espressione, nonché quelli relativi alla manipolazione intenzionale e spesso coordinata del servizio della piattaforma tramite creazione di account falsi, uso di programmi autonomi (*bot*) e altri comportamenti automatizzati (o parzialmente automatizzati) che possono condurre alla rapida e ampia diffusione di informazioni fraudolente»[175]. Le grandi piattaforme *online* sono tenute, infatti, a sottoporsi ad audit indipendenti al fine di verificare rischi "sistemici" dei loro algoritmi e dei relativi dati. Appare evidente che le riflessioni sulle criticità rilevate in dottrina valgono tanto più con riferimento all'applicazione dei sistemi algoritmici nel campo della medicina, considerate le conseguenze degli errori algoritmici in tale ambito. Non manca, infatti, chi auspica che la validazione delle tecnologie di IA nel campo medico andrà, in un futuro prossimo, a potenziare e integrarsi con le *skills* cliniche, diagnostiche e terapeutiche del medico e non a sovrapporsi o a sostituirsi a questi[176]. Per i profili relativi alla responsabilità dei danni causati da tali tecnologie, sarà (se non risolutiva) di grande ausilio la disciplina prevista dal legislatore europeo nell'*AI Act* e nella Direttiva sulla responsabilità extracontrattuale

[174] In particolare, cfr. art. 24, par. 2, e Considerando 77, art. 3, lett. m), p) e q), art. 33, *DSA*. Nel caso in cui il numero di utenti raggiunga oltre il 10% della popolazione dell'UE (ossia, 45 milioni di utenti) la Commissione potrà designare le piattaforme online molto grandi (*VLOPs*) o i motori di ricerca online molto grandi (*VLOSEs*), con la conseguente applicazione, per gli stessi, di obblighi aggiuntivi, come la valutazione del rischio e l'adozione di misure di mitigazione corrispondenti al livello di rischio rilevato.

[175] E. Garzonio, *L'algoritmo trasparente: obiettivi ed implicazioni della riforma dello Spazio digitale europeo*, in *Rivista italiana di informatica e diritto (RIID)*, 3, 2, 2021.

[176] A. Malva, V. Zurlo, *La medicina nell'era…*, *op. cit.*