

DANIELE CHIAPPINI^(*)

L'INTELLIGENZA ARTIFICIALE IN AMBITO MEDICO E LEGALE

ABSTRACT:The implementation of Artificial Intelligence (AI) in medical devices and its application in the legal field are topics of great significance and contemporary interest. This article thoroughly examines the utilization of AI in these domains, analyzing the pertinent European regulations, notably the Artificial Intelligence Act proposal and the European Regulation on Medical Devices. Throughout the analysis, the inquiry arises concerning how biases present in the training data impact the outcomes and potentially encroach upon fundamental rights, such as the right to health and a fair trial, thereby potentially discriminating against individuals belonging to underrepresented groups in these datasets. Employing an interdisciplinary approach, this study endeavors to identify solutions to address this issue by leveraging the existing regulations, which could serve as a foundation for their adaptation in other sectors beyond the ones under consideration.

SOMMARIO: 1. Introduzione. – 2. L'implementazione dell'IA nei dispositivi medici. – 3. L'utilizzo dei sistemi di IA in ambito giuridico. – 4. La costruzione della normativa europea sull'IA: l'*Artificial Intelligence Act*. – 5. Il regolamento europeo sui dispositivi medici. – 6. Problematiche comuni e risvolti etici. – 7. Pregiudizi e conseguenze in ambito medico e legale. – 8. Conclusioni.

1. — *Introduzione.*

L'Intelligenza Artificiale (IA) sta assumendo un ruolo sempre più rilevante in campo medico, offrendo numerose opportunità per migliorare la diagnosi e il trattamento delle malattie. Nell'ultimo decennio, ad esempio, la *Food and Drugs Administration* (FDA) ha esaminato e autorizzato la messa in commercio di un numero crescente di dispositivi che utilizzano tali funzioni in diversi campi della medicina e questa tendenza pare confermata anche

^(*) Consiglio Nazionale delle Ricerche – ISAFoM, Sezione distaccata di Perugia.

per il futuro⁽¹⁾. Tali sistemi sono di estrema rilevanza non solo per l'apporto diretto che forniscono ma anche perché producono informazioni ulteriori, cosa che ha condotto la FDA a rilasciare il sito web *Artificial Intelligence and Machine Learning*, mentre database simili sono state rilasciate anche da altri soggetti⁽²⁾⁽³⁾. Tuttavia, l'utilizzo di tali tecnologie richiede una regolamentazione adeguata al fine di garantire la sicurezza dei pazienti e prevenire rischi inerenti alla loro utilizzazione, a tal proposito, l'Unione Europea ha provveduto ad aggiornare la disciplina applicabile ai dispositivi medici che utilizzano l'Intelligenza Artificiale, oggi rinvenibile nel Regolamento sui Dispositivi Medici⁽⁴⁾ entrato in vigore il 26 maggio 2021, con un anno di ritardo rispetto a quanto previsto a causa della pandemia da COVID 19. Il regolamento impone requisiti rigorosi per la sicurezza e la qualità dei prodotti che utilizzano sistemi di IA e richiede un'ulteriore certificazione prima del loro utilizzo, ma non considera eventuali problematiche legate al bilanciamento dei dataset di addestramento.

Tale regolamentazione ha un impatto significativo anche sul piano giuridico, poiché si traduce in una maggiore responsabilità legale per i produttori e gli utilizzatori dei dispositivi medici contenenti IA. Permangono però ulteriori questioni che richiedono un'analisi accurata e una regolamentazione specifica. A tal proposito, la Commissione Europea ha presentato, il 21 aprile 2021, proposta di Regolamento sull'Intelligenza Artificiale l'*Artificial Intelligence Act*⁽⁵⁾ (AIA), prima proposta organica a livello internazionale volta ad armonizzare le normative che disciplinano questa tecnologia.

⁽¹⁾ www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices.

⁽²⁾ www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices.

⁽³⁾ [Aicentral.acrdsi.org](http://aicentral.acrdsi.org).

⁽⁴⁾ Regolamento (UE) 2017/745 del Parlamento Europeo e del Consiglio del 5 aprile 2017 relativo ai dispositivi medici, che modifica la direttiva 2001/83/CE, il regolamento (CE) n. 178/2002 e il regolamento (CE) n. 1223/2009 e che abroga le direttive 90/385/CEE e 93/42/CEE del Consiglio.

⁽⁵⁾ 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.

In tal modo i dispositivi medici che utilizzano o che sono basati su sistemi di Intelligenza Artificiale saranno soggetti tanto alle regole stabilite dalla normativa sui dispositivi medici che a quelle stabilite dal discutendo AIA.

Contemporaneamente, l'IA sta rivoluzionando il modo in cui vengono svolte numerose attività in ambito legale in tutto il mondo. In campo processuale, le tecnologie basate sull'IA possono supportare l'elaborazione di documenti legali, la ricerca di informazioni giuridiche e l'analisi dei dati relativi a casi legali. Ciò consente ai professionisti del diritto di effettuare ricerche più approfondite e di gestire un numero maggiore di casi in modo più efficiente. Inoltre, l'IA sta diventando sempre più importante anche per la polizia e l'amministrazione della giustizia. L'uso di tecnologie come i sistemi di riconoscimento facciale, la sorveglianza delle reti sociali e l'analisi dei dati biometrici, permette, nelle intenzioni, una maggiore efficacia nella prevenzione dei reati e nella conduzione delle attività processuali, tuttavia, il loro utilizzo solleva questioni importanti in merito alla privacy, alla protezione dei dati e alla possibilità di abusi, soprattutto a causa dei possibili errori commessi dai sistemi di IA.

In questo contesto, è essenziale comprendere l'impatto dell'IA a livello giuridico e legale. L'*Artificial Intelligence Act*, prende in considerazione anche i sistemi utilizzati nel campo giudiziario e di polizia e li categorizza allo stesso livello di quelli usati in ambito sanitario. Pertanto, l'obiettivo di questo articolo è di analizzare le implicazioni giuridiche della regolamentazione europea sull'utilizzo dell'IA nei dispositivi medici, dimostrando come le problematiche che emergono sono comuni anche ai sistemi di IA utilizzati in ambito legale e valutando se l'approccio utilizzato per l'approvazione dei sistemi di IA usati in medicina possa essere preso come esempio per quelli applicati al campo giuridico.

2. — *L'implementazione dell'IA nei dispositivi medici.*

Senza pretesa di voler essere esaustivi nell'elencare come i sistemi di Intelligenza Artificiale vengono utilizzati in ambito medico⁽⁶⁾, è sufficiente

⁽⁶⁾ Trovate un'analisi più nel dettaglio nell'articolo D. CHIAPPINI, *Legal and Ethical state of the art of Artificial Intelligence in Medicine*, in questa *Rivista*, 2020, pp. 97-142.

specificare che attualmente l'IA viene implementata in un gran numero di dispositivi medici e viene utilizzata in diversi campi della medicina moderna. Ad esempio, numerose IA vengono utilizzate per supportare la diagnosi di malattie, tra le quali quelle a cui viene prestata maggiormente attenzione sono il cancro, le malattie del sistema nervoso e le malattie cardiovascolari. Il progresso in tale materia ha anche la potenzialità di aiutare la diagnosi di malattie rare e lo sviluppo conseguente di terapie al fine di salvare e migliorare la vita dei pazienti che ne soffrono⁽⁷⁾. Un altro campo in cui l'Intelligenza Artificiale viene applicata dal punto di vista diagnostico è quello dell'imaging medico. È infatti possibile addestrare i sistemi di Intelligenza Artificiale a riconoscere con maggiore velocità e precisione⁽⁸⁾ i diversi pattern ottenuti tramite le varie tecniche di imaging radiologico come la tomografia computerizzata (TC), la risonanza magnetica (MRI), l'ecografia medica o gli ultrasuoni, la tomografia a emissione di positroni (PET) e la radiografia a raggi X. Un'ulteriore applicazione di tale tecnologia riguarda la creazione di farmaci. È infatti già stato creato un farmaco per trattare il disturbo ossessivo-compulsivo, la molecola DSP-1181, entrata nella sperimentazione clinica di Fase I in Giappone⁽⁹⁾. Lo sviluppo del farmaco ha richiesto un solo anno, contro i circa cinque anni di progetti simili con conseguente diminuzione dei costi di ricerca e altri farmaci creati con lo stesso procedimento supportato dall'IA sono in via di sperimentazione.

L'Intelligenza Artificiale è stata inoltre utilizzata anche nell'ambito della lotta al COVID-19. Sono stati creati sistemi di IA in grado di predire e tracciare i focolai e la diffusione del virus⁽¹⁰⁾, in grado di aiutare nella diagnosi del virus⁽¹¹⁾ e nello sviluppo dei vaccini⁽¹²⁾.

⁽⁷⁾ [Training.ace.bayer.com/en/pharma/artificial-intelligence-when-we-suddenly-know-what-we-dont-know](https://www.training.ace.bayer.com/en/pharma/artificial-intelligence-when-we-suddenly-know-what-we-dont-know).

⁽⁸⁾ K. CHOCKLEY, E. EMANUEL, *The End of Radiology? Three Threats to the Future Practice of Radiology*, in *Journal of the American College of Radiology*, 2016.

⁽⁹⁾ T. BURKI, *A new paradigm for drug development*, in *Lancet Digital Health*, 2020.

⁽¹⁰⁾ [Bluedot.global](https://bluedot.global).

⁽¹¹⁾ www.alizila.com/how-damo-academys-ai-system-detects-coronavirus-cases/.

⁽¹²⁾ www.coe.int/en/web/artificial-intelligence/ai-and-control-of-covid-19-coronavirus.

L'evoluzione tecnologica in materia ha portato numerose aziende ICT a sviluppare, da un lato, sistemi di IA da applicare al settore medico sanitario e, dall'altro, ad implementare sistemi di Intelligenza Artificiale in numerosi dispositivi che, pur non essendo qualificabili come dispositivi medici incrementano il numero di dati, anche sanitari, a disposizione, permettendo una maggiore efficienza degli output prodotti dai primi. Secondo «Artificial Intelligence in Medicine Market – Growth, Trends, and Forecast»⁽¹³⁾, questo porterà ad un incremento, nel solo settore medico, del valore di mercato dell'Intelligenza Artificiale dai 2,24 miliardi di dollari del 2018 ai 23,85 miliardi di dollari del 2025. Tra i prodotti di questo sistema ricordiamo Google Health⁽¹⁴⁾, un progetto volto a creare prodotti e servizi che connettono e danno significato alle informazioni sulla salute, IBM Watson Health, che intende integrare dati tecnologie e competenze per trasformare l'assistenza sanitaria⁽¹⁵⁾. Altre realtà hanno un approccio differente come quello di Apple, che si è specializzata nell'utilizzare i propri dispositivi per permettere una migliore comunicazione medico/paziente e tracciare lo stato di salute dei propri clienti.

3. — *L' utilizzo dei sistemi di IA in ambito giuridico.*

Il Parlamento Europeo ha osservato⁽¹⁶⁾ un crescente utilizzo dei sistemi di Intelligenza Artificiale in ambito giuridico da parte delle autorità giurisdizionali per diverse applicazioni. Tale impiego include l'uso di tecnologie di riconoscimento facciale e biometrico, la ricerca autonoma e l'analisi di database, la polizia predittiva⁽¹⁷⁾ e l'analisi delle scene del crimine. In

⁽¹³⁾ Research and markets, Artificial Intelligence in Medicine Market - Growth, Trends, and Forecast, reports 2019-2024 and 2020-2025.

⁽¹⁴⁾ *Health.google*.

⁽¹⁵⁾ *Www.ibm.com/it-it/watson-health*.

⁽¹⁶⁾ Parlamento Europeo, Risoluzione del 6 ottobre 2021 sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale.

⁽¹⁷⁾ Per polizia predittiva si intendono quell'insieme di attività finalizzate allo studio

aggiunta a queste applicazioni, vi sono servizi di risoluzione delle controversie online o sistemi che emettono decisioni automatizzate, ad esempio quelli adottati in Cina⁽¹⁸⁾, in Gran Bretagna⁽¹⁹⁾ e in Olanda⁽²⁰⁾. In Estonia⁽²¹⁾, il Ministero della Giustizia ha lanciato un programma pilota per automatizzare le sentenze dei tribunali per controversie di modesta entità riguardanti il pagamento di un massimo di € 7.000,00, appellabili presso un giudice umano. Inoltre, negli Stati Uniti d’America, l’utilizzo dell’IA in ambito giudiziario è già diffuso in numerose giurisdizioni⁽²²⁾ e, ad esempio, alcuni sistemi di Intelligenza Artificiale vengono già usati in ambito processuale per compiere una valutazione delle probabilità di recidiva degli imputati nonché quella di fuga di coloro che sono in attesa di processo⁽²³⁾. Tra i vari sistemi usati, un algoritmo della Arnold Foundation viene attualmente utilizzato in 21 giurisdizioni degli Stati Uniti per prevedere il comportamento degli imputati nella fase istruttoria⁽²⁴⁾, basandosi su un database di 1,5 milioni di procedimenti penali, mentre in Florida sistemi di IA vengono utilizzati per stabilire gli importi delle cauzioni⁽²⁵⁾.

e all’applicazione di metodi statistici che abbiano l’obiettivo di “prevedere” con un adeguato grado di certezza se un soggetto può commettere un reato, ma anche il luogo e quando questo potrebbe essere commesso. L’obiettivo di queste attività è la prevenzione delle attività criminali.

⁽¹⁸⁾ *English.bjinternetcourt.gov.cn/2019-07/01/c_190.htm*.

⁽¹⁹⁾ ONLINE DISPUTE RESOLUTION ADVISORY GROUP, *Online dispute resolution for low value civil claims*, 2015.

⁽²⁰⁾ Il sistema PA adottato in Olanda è in grado di emanare decisioni automatizzate che si fondano su autorizzazioni precedentemente concesse.

⁽²¹⁾ *Www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/*.

⁽²²⁾ A. ZAVRŠNIK, *Criminal justice, artificial intelligence systems, and human rights*, Era forum, 2020.

⁽²³⁾ A. ZAVRŠNIK, *Criminal justice, artificial intelligence systems, and human rights*, cit.

⁽²⁴⁾ D. CHIAPPINI, *Intelligenza Artificiale, un ritorno alla bouche de la loi?*, in *Rivista general de derecho procesal*, 2022, n. 58.

⁽²⁵⁾ L. ECKHOUSE, *Big data may be reinforcing racial bias in the criminal justice system*, in *Washington Post*, 2017.

Secondo un rapporto di *Markets and Markets*⁽²⁶⁾ pubblicato a gennaio 2021, il mercato globale dei sistemi di intelligenza artificiale per il settore legale è stato stimato a 1,07 miliardi di dollari nel 2020 e dovrebbe crescere a un tasso annuo composto del 31,3% nel periodo 2020-2026. Ciò significa che il mercato potrebbe raggiungere un valore di circa 4,31 miliardi di dollari entro il 2026.

Un primo elemento comune tra i due settori è rappresentato dall'efficienza dei responsi dell'IA rispetto a quelli delle controparti umane. Secondo i risultati di uno studio⁽²⁷⁾ condotto negli Stati Uniti su oltre un milione di casi di custodia cautelare, infatti, l'IA è già attualmente in grado prevedere con un grado di precisione superiore rispetto a un giudice umano se uno specifico imputato sarà propenso a tentare la fuga dalla libertà su parola nonché la sua possibilità di recidiva. Allo stesso modo, l'IA è più efficiente della controparte umana non solo nel settore legale, ma anche in ambito biomedico, dove è stato dimostrato⁽²⁸⁾ che i sistemi adottati per supportare i medici nella diagnosi di alcune patologie svolgono questo compito in modo più efficiente dei medici umani.

Questo crea conseguentemente le basi per valutare se sussistono problemi comuni correlati a tali risultati e se quindi possono essere risolti o migliorati attraverso un approccio regolamentare.

4. — *La costruzione della normativa europea sull'IA: l'Artificial Intelligence Act.*

Anche in considerazione della sua pervasività, tra le numerose sfide che l'IA presenta vi è anche quella relativa alla sua regolamentazione. Diversi

⁽²⁶⁾ MARKETS AND MARKETS, *Legal AI Software Market by Component (Solutions and Services), Deployment Mode, Technology, Application, End User (Corporate Legal Departments and Law Firms), and Region - Global Forecast to 2026*.

⁽²⁷⁾ J. KLEINBERG, H. LAKKARAJU, J. LESKOVEC, J. LUDWIG, S. MULLAINATHAN, *Human decisions and machine predictions*, in *The Quarterly Journal of Economics*, 2018, Volume 133, Issue 1, p. 237.

⁽²⁸⁾ Per maggiori approfondimenti vedi D. CHIAPPINI, *Legal and ethical state of the art of Artificial Intelligence in Medicine*, in questa *Rivista*, 2020.

paesi del mondo stanno infatti cercando di mettere in atto politiche ed approvare normative volte da un lato a disciplinare l'uso dell'IA e dall'altro a proteggere e tutelare i diritti e le libertà dei cittadini.

Tra questi, in Europa, la Commissione Europea ha presentato a fine aprile 2021 la proposta di regolamento sull'IA, che mira a creare un quadro normativo armonizzato per l'uso dell'IA in tutto il territorio dell'Unione e che è in grado di impattare anche al di fuori dei suoi confini (come accaduto già con il GDPR⁽²⁹⁾). La proposta prevede l'adozione di un approccio basato sul rischio che categorizza i sistemi di Intelligenza Artificiale in modo da prevedere diverse tipologie di regole a seconda della categoria di appartenenza.

Allargando il campo di studi oltre il panorama europeo, negli Stati Uniti si è passati da un'impostazione volta ad evitare una regolamentazione ad opera del legislatore⁽³⁰⁾ ad un approccio più simile a quello europeo, in cui l'uso dell'IA è regolamentato principalmente da leggi e regolamenti statali e federali⁽³¹⁾. Al riguardo, nel 2020, è stato emanato un memorandum sulla regolamentazione dell'IA⁽³²⁾. Anche in Asia molti paesi stanno adottando politiche e regolamenti per disciplinare l'uso dell'IA. In Cina, ad esempio, sono stati pubblicati numerosi documenti volti a regolamentare l'Intelligenza artificiale tra i quali spicca quello del 2017⁽³³⁾, che traccia l'approccio del paese allo sviluppo della tecnologia e delle applicazioni di IA, fissando obiettivi ampi fino al 2030.

In sintesi, la regolamentazione dell'IA rappresenta un tema cruciale a livello globale, capace di incidere anche sullo sviluppo della stessa e sulla sua diffusione.

⁽²⁹⁾ Reg. UE n. 2016/679.

⁽³⁰⁾ Ufficio esecutivo del Presidente degli Stati Uniti d'America, Consiglio Nazionale per la scienza e la tecnologia, Comitato sulla tecnologia, Prepararsi per il futuro dell'Intelligenza, 2016.

⁽³¹⁾ Congresso degli Stati Uniti, Testo - H.R.6216 - 116th (2019-2020): *National Artificial Intelligence Initiative Act of 2020*.

⁽³²⁾ UFFICIO ESECUTIVO DEL PRESIDENTE DEGLI STATI UNITI D'AMERICA, *Memorandum per i vertici dei dipartimenti esecutivi e agenzie. Guida alla regolamentazione delle applicazioni dell'Intelligenza Artificiale*, 17 novembre 2020.

⁽³³⁾ CONSIGLIO DI STATO CINESE, *Piano di sviluppo dell'IA di nuova generazione*, 2017.

Di particolare interesse è la costituenda normativa che l'Unione Europea sta realizzando, principalmente con l'*Artificial Intelligence Act*, che attualmente rappresenta la proposta di regolamentazione più avanzata a livello mondiale. Nella proposta di regolamento, si definisce come sistema di Intelligenza Artificiale «un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I e che può, per una data serie di obiettivi definiti dall'uomo, generare risultati quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono»⁽³⁴⁾. Trattandosi però di un atto *de iure condendo*, l'iter legislativo potrebbe portare ad una modifica di tale definizione, come suggerito nella proposta di compromesso del Consiglio del novembre 2022, che tende a limitare la definizione di cui all'articolo 3, paragrafo 1, ai sistemi sviluppati mediante approcci di apprendimento automatico e approcci basati sulla logica e sulla conoscenza⁽³⁵⁾, di cui viene data una definizione nei considerando⁽³⁶⁾.

L'*AI Act* non si limita solo a definire cosa sia un sistema di Intelligenza Artificiale, ma pone una categorizzazione usando un approccio basato sulla gestione del rischio. In tal modo suddivide i sistemi di IA in sistemi il cui uso è proibito, sistemi ad alto rischio e altri sistemi di IA, categoria comprendente sia i sistemi a rischio limitato che quelli a rischio minimo. L'appartenenza a tali categorie può mutare nel tempo al fine di garantire un corretto bilanciamento tra i diritti degli utenti e quelli dei produttori e, al contempo, non danneggiare il progresso scientifico nello sviluppo dell'IA. Per raggiungere

⁽³⁴⁾ 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, 2021, art. 3, par. 1, n. 1.

⁽³⁵⁾ La proposta del Consiglio prevede la seguente definizione di 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.

⁽³⁶⁾ Tali definizioni vengono introdotte nella proposta di compromesso del Consiglio Europeo nei considerando 6-bis e 6-ter.

tale obiettivo, è riconosciuto in capo alla Commissione il potere di adottare i necessari atti delegati al fine di includere e modificare gli allegati contenenti gli specifici sistemi di IA ad alto rischio, nonché i settori in cui vengono utilizzati. Viene altresì riconosciuta la possibilità di modificare i settori fondamentali ad alto rischio esistenti⁽³⁷⁾, prevedendo quindi una procedura snella e in grado di far fronte con maggior rapidità alle novità introdotte dallo sviluppo tecnologico che questa tecnologia può subire.

Il Titolo II dell'*Artificial Intelligence Act* tratta delle pratiche di Intelligenza Artificiale esplicitamente vietate⁽³⁸⁾ in quanto particolarmente dannose e in grado di contraddire i valori dell'Unione relativi al rispetto della dignità umana, della libertà, dell'uguaglianza, della democrazia, dello Stato di diritto e dei diritti fondamentali dell'Unione⁽³⁹⁾.

Maggiormente interessanti per il tema di cui trattiamo sono i sistemi ad alto rischio, nella cui categoria possono essere inclusi sia i sistemi utilizzati in ambito medico che giuridico. Appartengono a tale categoria, infatti, quei sistemi di IA che operano in modo autonomo e presentano un potenziale significativo di causare danni o pregiudizi a una o più persone. Tale potenziale dipende dal rapporto tra la gravità dei possibili danni e/o pregiudizi, dal grado di autonomia decisionale del sistema di IA, dalla probabilità di concretizzazione del rischio e dalle modalità e dal contesto in cui il sistema di IA viene utilizzato⁽⁴⁰⁾. Secondo quanto statuito nella proposta, questi sistemi devono

⁽³⁷⁾ Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale, Allegato alla risoluzione: Raccomandazioni dettagliate per l'elaborazione di un regolamento del Parlamento Europeo e del Consiglio sulla responsabilità per il funzionamento dei sistemi di intelligenza artificiale, art. 4, par. 2.

⁽³⁸⁾ 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, 2021, art. 5.

⁽³⁹⁾ 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, 2021, considerando 15.

⁽⁴⁰⁾ Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale, Allegato

essere sottoposti a una valutazione del rischio, i datasets utilizzati devono essere di alta qualità e i risultati devono essere tracciabili. Inoltre, devono essere rispettate le norme sulla documentazione tecnica, l'archiviazione dei dati e sulla trasparenza e deve essere prevista una supervisione umana e un alto livello di robustezza, sicurezza e accuratezza⁽⁴¹⁾.

Un sistema di Intelligenza Artificiale può essere qualificato ad alto rischio in due casi, in presenza dei quali il sistema potrà quindi essere incluso nell'elenco allegato al Regolamento. Il primo caso è rappresentato dalla situazione in cui siano soddisfatte contemporaneamente due condizioni⁽⁴²⁾. La prima è che il sistema sia destinato a essere utilizzato come componente di sicurezza di un prodotto, o sia esso stesso un prodotto, disciplinato dalla normativa di armonizzazione dell'Unione elencata nell'allegato II dell'atto. Tra gli atti inclusi in tale allegato rientrano il regolamento (UE) 2017/745 relativo ai dispositivi medici⁽⁴³⁾ e il regolamento (UE) 2017/746 relativo ai dispositivi medico-diagnostici in vitro⁽⁴⁴⁾.

La seconda condizione prevede che il prodotto il cui componente di sicurezza sia un sistema di intelligenza artificiale, o lo stesso sistema di IA sia il prodotto, debba essere sottoposto a una valutazione della conformità da parte di terzi ai fini dell'immissione sul mercato o della messa in servizio.

alla risoluzione: Raccomandazioni dettagliate per l'elaborazione di un regolamento del Parlamento Europeo e del Consiglio sulla responsabilità per il funzionamento dei sistemi di intelligenza artificiale, art. 3, lett. c).

⁽⁴¹⁾ 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, 2021, Titolo II.

⁽⁴²⁾ 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, 2021, art. 6.

⁽⁴³⁾ Regolamento (UE) 2017/745 del Parlamento europeo e del Consiglio, del 5 aprile 2017, relativo ai dispositivi medici, che modifica la direttiva 2001/83/CE, il regolamento (CE) n. 178/2002, il regolamento (CE) n. 1223/2009 e che abroga le direttive 90/385/CEE e 93/42/CEE del Consiglio (GU L 117 del 5 maggio 2017, p. 1).

⁽⁴⁴⁾ Regolamento (UE) 2017/746 del Parlamento europeo e del Consiglio, del 5 aprile 2017, relativo ai dispositivi medico-diagnostici in vitro e che abroga la direttiva 98/79/CE e la decisione 2010/227/UE della Commissione (GU L 117 del 5.5.2017, p. 176).

Sono inoltre considerati ad alto rischio i sistemi di IA inclusi nell'allegato III della risoluzione. Come anticipato, alla Commissione spetta il potere di adottare i necessari atti delegati al fine di modificare l'allegato secondo le necessità che potranno emergere nel corso del tempo.

La proposta di regolamento include nell'allegato III sistemi appartenenti a numerosi campi che saranno quindi considerati ad alto rischio, alcuni dei quali sono utilizzati in ambito giuridico. Rientrano infatti all'interno della categorizzazione proposta dalla Commissione quei sistemi destinati a essere utilizzati per l'identificazione biometrica remota "in tempo reale" e "a posteriori" delle persone fisiche⁽⁴⁵⁾, nonché i sistemi di IA destinati a essere utilizzati dalle autorità di contrasto per effettuare valutazioni individuali dei rischi delle persone fisiche al fine di determinare il rischio di reato o recidiva in relazione a una persona fisica o il rischio per vittime potenziali di reati. I sistemi di IA destinati a essere utilizzati come poligrafi e strumenti analoghi, o per rilevare lo stato emotivo di una persona fisica, quelli per individuare i "deep fake" e quelli destinati alla valutazione dell'affidabilità degli elementi probatori nel corso delle indagini o del perseguimento di reati. Sono altresì inclusi in questa categoria i sistemi di IA destinati a prevedere il verificarsi o il ripetersi di un reato effettivo o potenziale sulla base della profilazione delle persone fisiche di cui all'articolo 3, paragrafo 4, della direttiva (UE) 2016/680 o per valutare i tratti e le caratteristiche della personalità o il comportamento criminale pregresso di persone fisiche o gruppi, i sistemi volti alla profilazione delle persone fisiche di cui all'art. 3, par. 4 della direttiva (UE) 2016/680 nel corso delle indagini, dell'accertamento e del perseguimento di reati e i sistemi per l'analisi criminale riguardo alle persone fisiche, che consentono alle autorità di contrasto di eseguire ricerche in set di dati complessi, correlati e non correlati, resi disponibili da fonti di dati diverse o in formati diversi, al fine di individuare modelli sconosciuti o scoprire relazioni nascoste nei dati⁽⁴⁶⁾. Rientrano nella stessa categoria anche i sistemi

⁽⁴⁵⁾ Allegato III alla 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, n. 1.

⁽⁴⁶⁾ Allegato III alla proposta di regolamento del Parlamento Europeo e del Consiglio

utilizzati per la gestione della migrazione, dell'asilo e del controllo transfrontaliero⁽⁴⁷⁾, nonché, nell'ambito dell'amministrazione della giustizia, quelli destinati ad assistere l'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge⁽⁴⁸⁾.

Un maggiore approfondimento meritano i sistemi di identificazione biometrica remota, già estremamente utilizzati in ambito giuridico, e non, e che sono stati applicati, in paesi extra-UE come la Cina, anche come strumenti di controllo della popolazione. Ai sensi dello stesso regolamento⁽⁴⁹⁾, tali sistemi sono finalizzati all'identificazione a distanza di persone fisiche mediante il confronto dei dati biometrici di una persona con i dati biometrici contenuti in una banca dati di riferimento, e senza che l'utente del sistema di IA sappia in anticipo se la persona sarà presente e potrà essere identificata. Per sistema di identificazione biometrica in "tempo reale" invece si intende un sistema di identificazione biometrica remota in cui il rilevamento dei dati biometrici, il confronto e l'identificazione avvengono senza ritardi significativi. Sono ricomprese, oltre alle identificazioni istantanee, anche quelle che avvengono con brevi ritardi, al fine di evitare l'elusione della normativa. A tali sistemi si contrappongono quelli la cui identificazione biometrica remota avviene a posteriori, definiti in via residuale rispetto alla categoria precedente. Per spazi accessibili al pubblico si intende un qualsiasi luogo fisico accessibile al pubblico, indipendentemente dall'applicabilità di specifiche condizioni di accesso.

Per i sistemi di identificazione biometrica "in tempo reale" in spazi ac-

che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'unione, n. 6.

⁽⁴⁷⁾ Allegato III alla 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, n. 7.

⁽⁴⁸⁾ Allegato III alla 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, n. 8.

⁽⁴⁹⁾ 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, 2021, art. 3, nn. 35 ss.

cessibili al pubblico è previsto un divieto specifico di utilizzo⁽⁵⁰⁾, che viene meno qualora risulti strettamente necessario per la ricerca mirata di potenziali vittime di crimini, ivi inclusi i bambini scomparsi, prevenire una minaccia specifica sostanziale e imminente per la vita o l'incolumità delle persone o un attacco terroristico. È altresì prevista la facoltà di utilizzare tali sistemi al fine di prevenire una specifica sostanziale e imminente minaccia alla vita o un attacco terroristico, oppure per permettere l'individuazione, la localizzazione, l'identificazione o il perseguimento di un autore o sospettato di un reato di cui all'art. 2, par. 2, della decisione quadro del Consiglio 2002/584/GAI⁽⁵¹⁾ punibile con una pena detentiva avente un massimo edittale di almeno tre anni secondo la legislazione di tale Stato.

Tali deroghe sono soggette al rispetto di specifici requisiti⁽⁵²⁾, tra i quali spicca la concessione di un'autorizzazione preventiva, rilasciata conseguentemente alla presentazione di una richiesta motivata da un'autorità giudiziaria o amministrativa indipendente dello Stato membro in cui è richiesto

⁽⁵⁰⁾ 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, 2021, art. 5, par. 1, lett. *d*).

⁽⁵¹⁾ La lista dei reati include partecipazione a un'organizzazione criminale, terrorismo, tratta di esseri umani, sfruttamento sessuale dei bambini e pornografia infantile, traffico illecito di stupefacenti e sostanze psicotrope, traffico illecito di armi, munizioni ed esplosivi, corruzione, frode, riciclaggio di proventi di reato, falsificazione di monete, criminalità informatica, criminalità ambientale, compreso il traffico illecito di specie animali protette e il traffico illecito di specie e di essenze vegetali protette, favoreggiamento dell'ingresso e del soggiorno illegali, omicidio volontario, lesioni personali gravi, traffico illecito di organi e tessuti umani, rapimento, sequestro e presa di ostaggi, razzismo e xenofobia, furti organizzati o con l'uso di armi, traffico illecito di beni culturali, compresi gli oggetti d'antiquariato e le opere d'arte, truffa, racket ed estorsioni, contraffazione e pirateria in materia di prodotti, falsificazione di atti amministrativi e traffico di documenti falsi, falsificazione di mezzi di pagamento, traffico illecito di sostanze ormonali ed altri fattori di crescita, traffico illecito di materie nucleari e radioattive, traffico di veicoli rubati, stupro, incendio volontario, reati che rientrano nella competenza giurisdizionale della Corte penale internazionale, dirottamento di aereo/nave, sabotaggio.

⁽⁵²⁾ 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, 2021, art. 5, par. 3.

l'utilizzo. La suddetta autorizzazione può essere accordata solo in presenza di prove oggettive o indicazioni chiare dalle quali emerga che l'utilizzo del sistema è proporzionato rispetto al conseguimento degli obiettivi per i quali è prevista la deroga. Deve essere appurato che la gravità, probabilità ed entità del danno causato dal mancato uso del sistema sia tale da rendere l'utilizzo in questione necessario, anche in considerazione delle possibili conseguenze riguardanti i diritti e le libertà di tutte le persone interessate. È in ogni caso prevista la possibilità, in presenza di una situazione di urgenza adeguatamente giustificata, di proporre la richiesta di autorizzazione all'utilizzo del sistema solo durante o dopo l'uso dello stesso.

La proposta prevede che Stato membro abbia la possibilità di autorizzare in tutto o in parte l'uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto, entro i limiti precedentemente descritti. Qualora uno stato membro volesse avvalersi di tali possibilità, dovranno essere stabilite tutte le regole necessarie all'emanazione della richiesta, al suo rilascio, all'esercizio delle autorizzazioni e alle relative attività di controllo. Al riguardo, il testo di compromesso del Consiglio chiarisce quali sono le situazioni nelle quali tale uso è considerato strettamente necessario a fini di contrasto e quali sono quelle in cui l'autorizzazione all'utilizzo di tali sistemi dovrebbe essere considerata una misura eccezionale.

La proposta così illustrata però non collima con la posizione del Parlamento Europeo⁽⁵³⁾, più volte reiterata, che ha proposto che venga applicato un divieto permanente dell'utilizzo dei sistemi di analisi e/o riconoscimento automatici negli spazi pubblici⁽⁵⁴⁾. Alla base di questa posizione vi sono gli errori di identificazione e di classificazione commessi dagli algoritmi attualmente in uso, errori che sono in grado di incidere negativamente sui diritti

⁽⁵³⁾ Risoluzione del Parlamento europeo del 6 ottobre 2021 sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale, art. 9.

⁽⁵⁴⁾ Risoluzione del Parlamento europeo del 6 ottobre 2021 sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale, art. 26.

di specifici gruppi⁽⁵⁵⁾. Allo stesso modo l'eurocamera ha anche richiesto che venga approvata una moratoria sulla diffusione dei sistemi di riconoscimento facciale per le attività di contrasto con funzione di identificazione, salvo che le attività non siano finalizzate all'identificazione delle vittime di reati, fintanto che il livello tecnico dei sistemi di IA non possa essere considerato rispettoso dei diritti fondamentali e gli output di tali sistemi siano privi di distorsioni e non discriminatori⁽⁵⁶⁾. Secondo quanto emerso finora, la stessa posizione verrà quindi tenuta dall'eurocamera in sede di approvazione del testo dell'AI Act, come confermato dall'ultimo voto delle commissioni del Parlamento Europeo che si sono espresse in materia⁽⁵⁷⁾.

Nella proposta di compromesso del Consiglio è previsto un ulteriore livello orizzontale con l'obiettivo di garantire che non siano inclusi nella categoria in oggetto quei sistemi di IA che non presentano effettivamente il rischio di causare gravi violazioni dei diritti fondamentali o altri rischi significativi. Nella controproposta, sono indicate delle disposizioni⁽⁵⁸⁾ in base alle quali i sistemi di IA contenuti nell'allegato III non saranno considerati ad alto rischio qualora l'output del sistema sia puramente accessorio in relazione all'azione o alla decisione da adottare non comportando quindi un rischio significativo per la salute, per la sicurezza o per i diritti fondamentali.

⁽⁵⁵⁾ In particolare, l'uso di tali sistemi viene ritenuto in grado di danneggiare i soggetti appartenenti a determinati gruppi razziali o etnici, le persone LGBTI, i bambini, gli anziani e le donne. Viene anche ricordato il diritto degli individui non solo ad essere identificati correttamente ma anche a non essere identificati affatto.

⁽⁵⁶⁾ Risoluzione del Parlamento europeo del 6 ottobre 2021 sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale, art. 27.

⁽⁵⁷⁾ L'approvazione è avvenuta l'11 maggio 2023 nella riunione congiunta delle commissioni per il Mercato interno e la protezione dei consumatori (IMCO) e per le Libertà civili, la giustizia e gli affari interni (LIBE) del Parlamento UE.

⁽⁵⁸⁾ Proposta di compromesso del Consiglio Europeo alla 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, 2022, art. 6, par. 3.

Nella proposta in oggetto sono poi state aggiunte regole specifiche per i sistemi di Intelligenza Artificiale per finalità generali. Questi sono definiti come sistemi di IA che, indipendentemente dalla modalità con cui sono immessi sul mercato o messi in servizio, sono destinati, nelle intenzioni del fornitore, a svolgere funzioni di applicazione generale quali il riconoscimento di immagini o voce, la creazione di audio o video, la rilevazione di modelli, la risposta a domande, la traduzione o altro. In tal modo si prendono in considerazione tutte quelle situazioni in cui dei sistemi di IA potrebbero essere utilizzati per molteplici scopi e conseguentemente, laddove integrati in un altro sistema, potrebbero comportare rischi significativi, diventando conseguentemente ad alto rischio. In tal modo, un sistema di IA per finalità generali può essere utilizzato in una varietà di contesti e integrato all'interno di diversi sistemi di IA⁽⁵⁹⁾.

La proposta di compromesso suggerisce quindi che i requisiti richiesti per i sistemi di Intelligenza Artificiale ad alto rischio si debbano applicare anche ai sistemi di IA per finalità generali nel caso in cui questi vengano utilizzati come sistemi ad alto rischio o come componenti di sistemi ad alto rischio. Affinché ciò sia possibile sarà però necessario determinare le modalità di tale applicazione, e ciò potrà avvenire conseguentemente all'adozione di un atto di esecuzione da parte della Commissione Europea.

Per quello che riguarda gli altri sistemi di IA, ovvero quelli che non vengono qualificati come sistemi ad alto rischio, questi ricomprendono sia quelli a rischio limitato⁽⁶⁰⁾ che quelli a rischio minimo⁽⁶¹⁾. I primi sono quei sistemi che presentano un chiaro rischio di manipolazione (come, ad esempio, *chatbot* o *deepfake*) in relazione al quale gli utenti dovranno essere informati

⁽⁵⁹⁾ Proposta di compromesso del Consiglio Europeo alla 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, 2022, art. 3, par. 1-*ter*.

⁽⁶⁰⁾ *New rules for Artificial Intelligence. Questions and Answers*, in ec.europa.eu/commission/press-corner/detail/en/QANDA_21_1683.

⁽⁶¹⁾ *New rules for Artificial Intelligence. Questions and Answers*, in ec.europa.eu/commission/press-corner/detail/en/QANDA_21_1683.

del fatto che non stanno interagendo con un essere umano. A tali sistemi, nell'*Artificial Intelligence Act* vengono quindi imposti requisiti di trasparenza specifici. Per sistemi a rischio minimo si intendono invece tutti gli altri sistemi di IA che non sono ricompresi nelle precedenti categorie. Tali sistemi possono essere sviluppati e utilizzati in base alla legislazione esistente senza che siano previsti ed imposti ulteriori obblighi legali.

Sulla base di tali dati è possibile affermare che i sistemi di Intelligenza Artificiale utilizzati in ambito legale e medico apparterranno, una volta approvato il regolamento, in via principale, ma non esclusiva, alla categoria dei sistemi ad alto rischio.

5. — *Il regolamento Europeo sui dispositivi medici.*

Al fine di esaurire la panoramica relativa alla regolamentazione dei sistemi di Intelligenza artificiale, è bene trattare brevemente del Regolamento sui dispositivi medici (MDR)⁽⁶²⁾ recentemente entrato in vigore. Nonostante non esistano normative specifiche dedicate all'IA, tale regolamento considera e regola l'utilizzo dei sistemi di Intelligenza Artificiale all'interno dei dispositivi medici. Il regolamento doveva entrare in vigore il 26 maggio 2020, ma, a causa della pandemia di COVID-19, l'entrata in vigore è stata posticipata di un anno. Il nuovo Regolamento abroga così la direttiva sui dispositivi medici (93/42/CEE – MDD) e la Direttiva sui dispositivi medici impiantabili attivi (90/385/EEC – AIMD)⁽⁶³⁾.

Il nuovo MDR apporta alcune modifiche alla classificazione dei dispositivi medici, includendo nella definizione dei dispositivi medici alcuni software che, secondo la precedente direttiva, non erano inclusi. Di conseguenza, la nuova definizione di dispositivo medico presentata nell'art. 2 (1) comprende

⁽⁶²⁾ Regolamento (UE) 2017/45 del Parlamento Europeo e del Consiglio del 5 aprile 2017 relativo ai dispositivi medici, che modifica la direttiva 2001/83/CE, il regolamento (CE) n. 178/2002 e il regolamento (CE) n. 1223/2009 e che abroga le direttive 90/385/CEE e 93/42/CEE del Consiglio.

⁽⁶³⁾ Art. 122 of the Medical Devices Regulation (2017/745).

il software utilizzato allo scopo medico di previsione o prognosi di malattia come dispositivo medico. In ogni caso il Regolamento compie un'importante distinzione, prevedendo che «il software specificamente destinato dal fabbricante a essere impiegato per una o più delle destinazioni d'uso mediche indicate nella definizione di dispositivo medico si considera un dispositivo medico, mentre il software destinato a finalità generali, anche se utilizzato in un contesto sanitario, o il software per fini associati allo stile di vita e al benessere non è un dispositivo medico»⁽⁶⁴⁾.

Analogamente alla Direttiva abrogata, l'MDR classifica i dispositivi medici in quattro categorie (Classi I, II a, II b, e III) in base alla destinazione d'uso dei dispositivi medici e ai loro rischi intrinseci. Il regolamento introduce nuove regole di classificazione esplicitamente incentrate sul software. Secondo quanto previsto nell'allegato VIII, il software destinato a fornire informazioni utilizzate per prendere decisioni a fini diagnostici o terapeutici rientra nella classe IIa, a meno che tali decisioni abbiano effetti tali da poter causare il decesso o un deterioramento irreversibile delle condizioni di salute di una persona, nel qual caso rientra nella classe III. Qualora invece possano provocare un grave deterioramento delle condizioni di salute di una persona rientrano nella classe IIb. Il software destinato a monitorare i processi fisiologici rientra nella classe IIa, a meno che sia destinato a monitorare parametri fisiologici vitali e la natura delle variazioni di detti parametri sia tale da poter creare un pericolo immediato per il paziente, nel qual caso rientra nella classe IIb. Tutti gli altri software rientrano nella classe I⁽⁶⁵⁾.

Queste nuove regole comportano una riclassificazione del software e, di conseguenza, nell'ottobre 2019, il Gruppo di coordinamento dei dispositivi medici ha rilasciato una guida non vincolante sulla qualificazione e classificazione del software ai sensi del regolamento⁽⁶⁶⁾.

⁽⁶⁴⁾ Regolamento (EU) 2017/745 sui Dispositivi Medici, considerando 19.

⁽⁶⁵⁾ Regolamento (EU) 2017/745 sui Dispositivi Medici, Allegato VIII, Capo III, Regola 11.

⁽⁶⁶⁾ Medical Device coordination group document, MDCG 2019-11, *Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR*, 2019.

Il nuovo Regolamento deve essere accompagnato da una marcatura CE che attesti la conformità del dispositivo ai requisiti previsti dal MDR affinché possa circolare liberamente all'interno dell'UE⁽⁶⁷⁾. In particolare, i fabbricanti di dispositivi medici devono effettuare una valutazione della conformità dei loro dispositivi prima della loro immissione sul mercato⁽⁶⁸⁾. Questa valutazione deve essere basata sulla classificazione e sul tipo del particolare dispositivo⁽⁶⁹⁾. Ad esempio⁽⁷⁰⁾, i dispositivi di classe I hanno un basso livello di vulnerabilità e quindi la procedura di valutazione della conformità può essere generalmente eseguita sotto la sola responsabilità dei produttori⁽⁷¹⁾. Al contrario, i dispositivi di classe II a, II b e III che presentano un rischio più elevato rispetto ai dispositivi di classe I comportano il coinvolgimento di un organismo notificato, un organismo di valutazione della conformità designato secondo l'MDR⁽⁷²⁾.

Secondo quanto previsto dal discutendo *Artificial Intelligence Act*, i sistemi di intelligenza artificiale considerati in tale regolamento apparterranno alla categoria dei sistemi di IA ad alto rischio in quanto potenzialmente in grado di creare danni o pregiudizi alla salute.

6. — *Problematiche comuni e risvolti etici.*

Quanto esplicitato finora da un lato fornisce una panoramica sugli utilizzi dell'IA nei campi medico e giuridico e dall'altro sulla regolamentazione

⁽⁶⁷⁾ Regolamento (EU) 2017/745 sui Dispositivi Medici, Considerando 40, Regolamento (EU) 2017/745 sui Dispositivi Medici, art. 2 (43).

⁽⁶⁸⁾ Regolamento (EU) 2017/745 sui Dispositivi Medici, art. 52, Regolamento (EU) 2017/745 sui Dispositivi Medici, Allegati IX-XI.

⁽⁶⁹⁾ Regolamento (EU) 2017/745 sui Dispositivi Medici, art. 52.

⁽⁷⁰⁾ S. GERKE, T. MINNSEN, G. COHEN, *Ethical and legal challenges of artificial intelligence-driven healthcare*, in *Artificial Intelligence in Healthcare*, 2020, pp. 295-336.

⁽⁷¹⁾ Regolamento (EU) 2017/745 sui Dispositivi Medici, Considerando 60, Regolamento (EU) 2017/745 sui Dispositivi Medici, art. 52 (7).

⁽⁷²⁾ Regolamento (EU) 2017/745 sui Dispositivi Medici, Considerando 60, Regolamento (EU) 2017/745 sui Dispositivi Medici, art. 2 (42).

che l'Unione Europea sta implementando per far sì che l'utilizzo di questa tecnologia possa essere il più sicuro, robusto e adeguato ma contemporaneamente non impatti negativamente sulla ricerca condotta in questo ambito.

Eppure, l'utilizzo dell'IA oltre ad impattare sul mondo del lavoro, appare capace di incidere direttamente sui diritti fondamentali quali quello alla salute (art. 32 Cost.), quello ad un giusto processo (art. 111 Cost.) e in generale sul principio di uguaglianza (art. 3 Cost.). Questo fa emergere una serie di problematiche etico-giuridiche.

Tra questi problemi possiamo rilevare che l'accesso universale all'intelligenza artificiale in campo medico e un accesso equo e paritario all'assistenza sanitaria che utilizzi l'IA rappresentano un discrimine estremamente importante già in questo momento e lo rappresenteranno ancora di più in futuro. Lo stesso ragionamento può essere condotto per l'utilizzo dell'IA in ambito giuridico. Tale problema sussiste non solamente per coloro vivono nei paesi in via di sviluppo, ma è presente anche nei paesi industrializzati, in relazione quanti, a causa di particolari condizioni di vita o reddito, potrebbero non avere accesso a tutto ciò che la moderna tecnologia mette a disposizione. Inoltre, interi gruppi potrebbero avere accesso maggiormente difficoltoso. Se tale problema può apparire di facile comprensione dal punto di vista medico, in quanto già oggi si assiste a limitate possibilità di accesso a strutture o terapie sulla base dell'appartenenza ad uno specifico Paese o alla disponibilità economica, anche in ambito giuridico, ove il principio di uguaglianza sembrerebbe essere maggiormente rispettato, si possono notare, come sarà esposto più avanti, problemi simili.

Affrontando la questione dell'uso dell'IA in ambito medico, il Consiglio d'Europa ha affermato⁽⁷³⁾ che anche in circostanze eccezionali la qualità dei servizi ospedalieri e delle strutture sanitarie non dovrebbe essere ridotta né compromettere l'accesso universale alle cure, e il supporto fornito dalla tecnologia dell'informazione e dall'intelligenza artificiale nel settore sanitario non dovrebbe essere considerato solo come un modo per ridurre i costi, ri-

⁽⁷³⁾ CONSIGLIO D'EUROPA, *AI and control of Covid-19 coronavirus*, in www.coe.int/en/web/artificial-intelligence/ai-and-control-of-covid-19-coronavirus.

tenendo necessario considerare i problemi che potrebbero sorgere in termini di qualità dei servizi e accesso globale all'assistenza sanitaria e garantendo quindi la stessa, comprensiva dell'uso dell'IA, a tutti.

Da un punto di vista dell'uso dell'IA in ambito giuridico e di governance, è stata la pandemia da coronavirus a dimostrare come l'uso di tale tecnologia può impattare negativamente sui diritti fondamentali, come spiegato nelle linee guida etiche dell'Unione Europea sull'intelligenza artificiale affidabile⁽⁷⁴⁾. Infatti, conseguentemente alla diffusione della pandemia alcuni paesi hanno adottato misure per controllare e monitorare in massa la popolazione tramite questa tecnologia, approntando protocolli che poi sono stati affinati e utilizzati anche al di là dell'obiettivo contingente del contenimento sanitario.

Questi aspetti riguardano tanto l'ambito giuridico quanto quello sanitario in quanto utilizzano dati biometrici, dati di geolocalizzazione, riconoscimento facciale o dati sanitari, per i quali andrebbe garantito un trattamento con elevati standard di protezione. Inoltre l'adozione di misure specifiche, sia in situazioni di emergenza che ordinarie, dovrebbe avvenire nel rispetto della dignità e della vita privata degli utenti, considerando i diversi possibili pregiudizi che potrebbero emergere attraverso l'uso dei vari tipi di operazioni di sorveglianza, in quanto capaci di causare discriminazioni significative.

Un ulteriore ambito di contatto tra le due materie è quello che coinvolge le motivazioni che sono alla base degli output dei sistemi di IA interrogati da medici e giuristi e che quindi sono anche alla base delle decisioni prese da questi ultimi.

Le applicazioni di intelligenza artificiale utilizzate in ambito sanitario sono in grado di trasformare la relazione tra medico e paziente, incidendo sul principio del consenso informato. Appare infatti necessario indagare in quali circostanze si dovrebbero applicare i principi del consenso informato quando le procedure cliniche coinvolgono l'intelligenza artificiale, ed è particolarmente importante stabilire fino a che punto i medici abbiano la responsabilità di spiegare al paziente le complessità dell'intelligenza artificiale

⁽⁷⁴⁾ GRUPPO INDIPENDENTE DI ESPERTI DI ALTO LIVELLO SULL'INTELLIGENZA ARTIFICIALE, *Orientamenti etici per un'IA affidabile. Requisiti di un'IA affidabile*, 2019, p. 15.

e il suo utilizzo nelle procedure. Il problema principale, da questo punto di vista, è rappresentato dalla conoscenza che ha il medico del funzionamento del sistema utilizzato, e quindi la sua competenza nel definire e spiegare i diversi tipi di apprendimento automatico utilizzati dall'IA, quali sono i dati utilizzati come input e per l'addestramento, nonché la valutazione circa la possibile presenza di pregiudizi presenti. A proposito di questo, è stato osservato che i pazienti generalmente non sono consapevoli se i loro medici utilizzano strumenti decisionali computerizzati per definire il trattamento e ricevono poche informazioni su quali fonti vengono utilizzate dai loro medici per pervenire ad una decisione⁽⁷⁵⁾. Trovare una soluzione a questa problematica è particolarmente difficile laddove il sistema di Intelligenza Artificiale è così avanzato e complesso da comportare un'impossibilità pressoché totale nel definire le motivazioni alla base dell'output. In tali casi si parla di utilizzo di algoritmi a "scatola nera" (o *black box*), che si basano su tecniche di apprendimento automatico non conoscibili o su relazioni troppo complesse per essere comprese esplicitamente⁽⁷⁶⁾ e che possono portare a cure sub-ottimali per i pazienti⁽⁷⁷⁾. In questo caso, sarebbe necessario stabilire fino a che punto un medico abbia l'obbligo di dichiarare al paziente di non essere in grado di comprendere appieno i processi che stanno alla base della diagnosi/trattamento consigliato dall'IA.

Questo aspetto è, inoltre, profondamente connesso alla tutela dei dati personali, in quanto secondo quanto stabilito dall'art. 22 GDPR, l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente

⁽⁷⁵⁾ I.G. COHEN, R. AMARASINGHAM, A. SHAH, B. XIE, B. LO, *The legal and ethical concerns that arise from using complex predictive analytics in health care*, in *Health Affairs*, 2014, pp. 1139-1147.

⁽⁷⁶⁾ hls.harvard.edu/today/petrie-flom-center-launches-project-precision-medicine-artificial-intelligence-law-pmail/, ultima visita maggio 2023.

⁽⁷⁷⁾ Un tipico esempio di "scatola nera" è rappresentato da quei sistemi di Intelligenza Artificiale vengono utilizzati per individuare arresti cardiaci durante le chiamate di emergenza utilizzando "algoritmi di Corti" il cui stesso inventore non sa in che modo il software giunge alle sue decisioni per avvisare gli operatori di emergenza che qualcuno ha un arresto cardiaco. L'assenza di una spiegabilità e consapevolezza può quindi produrre errori e danneggiare i pazienti.

sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona.

Parrebbe lecito però domandarsi in che modo una questione così intimamente connessa al campo medico quale il rapporto medico-paziente e il consenso informato possa essere traslabile in ambito legale. Eppure, la risposta è facilmente comprensibile, se si pensa da un lato al rapporto tra avvocato ed assistito. Anche in tal caso si viene a strutturare un rapporto fiduciario e l'art. 27 del codice deontologico forense stabilisce che l'avvocato deve informare chiaramente la parte assistita, all'atto dell'assunzione dell'incarico, delle caratteristiche e dell'importanza di quest'ultimo e delle attività da espletare, precisando le iniziative e le ipotesi di soluzione. L'elevato impatto che l'IA sta avendo anche nel settore giuridico è dimostrato anche dal fatto che un sistema di Intelligenza Artificiale, agendo come avvocato virtuale, ha svolto un ruolo attivo e diretto per la prima volta in un procedimento penale⁽⁷⁸⁾. Tale circostanza si è verificata in un tribunale inglese, dove, nel corso di un processo, un Intelligenza Artificiale sviluppata dalla startup "DoNotPay" ha interagito direttamente con l'imputato, fornendo precise istruzioni tramite un dispositivo mobile e auricolari.

L'impatto maggiore però in ambito legale lo si ha nel supporto che viene fornito al giudice. Anche in questo caso emerge il problema della scatola nera in relazione all'obbligo di motivazione dei provvedimenti giurisdizionali⁽⁷⁹⁾, obbligo posto al fine di assicurare concretamente il perseguimento dei principi costituzionali posti in tema di giurisdizione, come ad esempio il diritto di difesa, il principio di indipendenza del giudice, la soggezione dello stesso alla legge, nonché per assicurare il rispetto del principio di legalità. In che modo si rapporta l'obbligo di motivazione, e la decisione sottesa alla stessa con l'ausilio fornito da un Intelligenza Artificiale? Se da un lato la risposta semplice potrebbe essere che l'IA non è altro che il corrispettivo artificiale di un Con-

⁽⁷⁸⁾ A. MARTINEZ, *La professione di avvocato e l'intelligenza Artificiale*, in *Sole24ore*, 1° febbraio 2023.

⁽⁷⁹⁾ Costituzione italiana, art. 111, 6° comma: «Tutti i provvedimenti giurisdizionali devono essere motivati».

sulente Tecnico d'Ufficio (CTU), la realtà è più complessa, in quanto il CTU può sempre essere interrogato più approfonditamente per comprendere le ragioni alla base della perizia, mentre, in presenza del problema della scatola nera, l'IA non offre lo stesso grado di trasparenza. In tal modo si potrebbe venire a creare un vulnus derivante dal fatto che il giudice ha sì a disposizione una valutazione compiuta da un sistema di IA che offre una soluzione che può apparire coerente e circostanziata, ma la conseguente motivazione del provvedimento può risultare carente dal punto di vista sostanziale in quanto, sebbene metodologicamente accurata, la motivazione in questione potrebbe ridursi ad una mera accettazione del responso fornito dal sistema interrogato o alla ricerca di ulteriori conferme di detto responso. Questo non solo per semplicità d'uso, ma anche perché discostarsi da un "parere" fornito da un sistema costruito per svolgere quella specifica funzione può risultare difficile. In tal modo si verrebbe anche a creare un sistema ricorsivo in base al quale si rischia di avere un'uniformazione delle sentenze e decisioni in quanto il sistema di IA interrogato vedendo incrementare un determinato tipo di decisioni le assumerà in modo sempre più costante come base per il proprio output.

In tal modo si può venire a creare una situazione di affidamento da parte del giudice, soprattutto per questioni particolarmente discrezionali, in grado di incidere non solo sull'aspetto della motivazione ma direttamente sulla vita delle persone. Riprendendo quanto accennato in precedenza (cfr. par. 3), i sistemi di IA sono ormai utilizzati in numerose giurisdizioni. Prendendo come esempio l'uso di tali sistemi negli Stati Uniti⁽⁸⁰⁾, all'avanguardia da questo punto di vista, è stato rilevato come, similmente a quanto accade in campo medico⁽⁸¹⁾, l'IA è in grado di prevedere con maggior precisione rispetto a un giudice umano il comportamento di un soggetto⁽⁸²⁾. Tutto ciò potrebbe quindi far presupporre che la questione sia meramente accademica-

⁽⁸⁰⁾ Ma la questione è valida sia in campo civile che penale. La scelta dell'utilizzo del campo penale è determinata dal fatto che vi sono un numero maggiore di studi e dati utilizzabili.

⁽⁸¹⁾ K. CHOCKLEY, E. EMANUEL, *The End of Radiology? Three Threats to the Future Practice of Radiology*, in *Journal of the American College of Radiology*, 2016.

⁽⁸²⁾ J. KLEINBERG, H. LAKKARAJU, J. LESKOVEC, J. LUDWIG, S. MULLAINATHAN, *Human decisions and machine predictions*, in *The Quarterly Journal of Economics*, 2018, Volume 133, Issue 1, p. 237.

ca, e che la necessità di motivazione possa essere ritenuta soddisfatta con il semplice riferimento al sistema utilizzato.

La realtà però è più complessa, poiché tale posizione non tiene in considerazione un'ulteriore problematica, quella relativa alla possibilità di *bias*, o pregiudizi, insiti nel sistema di IA. Sebbene infatti possa apparire strano, un sistema di Intelligenza Artificiale è scevro da pregiudizi, corretto e in grado di fornire un output, o decisione, “giusta” solamente se i dataset con i quali è stato addestrato lo sono⁽⁸³⁾. Tali dataset sono delle raccolte di dati che vengono utilizzate per insegnare all'IA a riconoscere modelli, apprendere regole e fare previsioni e possono comprendere testi, immagini, video, audio o altri tipi di dati, a seconda del tipo di problema che dovrà essere risolto.

I dataset di addestramento sono curati da esperti umani, che, nel caso dell'apprendimento supervisionato⁽⁸⁴⁾, forniscono etichette o marcatori per indicare la corretta risposta o classificazione per ciascun esempio presente nel dataset. È particolarmente importante che i dataset utilizzati per l'addestramento siano rappresentativi della “popolazione” considerata e diversificati per garantire che l'IA sia in grado di compiere una generalizzazione corretta e di adattarsi a nuove situazioni. Maggiore è il volume dei dati disponibili e più corretto sarà l'output, in quanto l'IA trae informazioni e apprende dai modelli presenti nel dataset. Conseguentemente, dati maggiormente completi e rappresentativi comporteranno migliori prestazioni e maggiore efficacia nell'affrontare scenari inediti e nuovi compiti.

⁽⁸³⁾ GRUPPO INDIPENDENTE DI ESPERTI DI ALTO LIVELLO SULL'INTELLIGENZA ARTIFICIALE, *Orientamenti etici per un'IA affidabile. Requisiti di un'IA affidabile*, cit., p. 15.

⁽⁸⁴⁾ L'apprendimento supervisionato è un tipo di approccio in cui un algoritmo viene addestrato utilizzando un dataset di addestramento che comprende sia le caratteristiche (*input*) che le corrispondenti etichette o risultati desiderati (*output*). L'obiettivo dell'apprendimento supervisionato è quello di addestrare l'algoritmo in modo che possa imparare a fare previsioni o prendere decisioni basate sui dati di input forniti, al fine di ottenere l'output desiderato.

Durante il processo di addestramento, l'algoritmo analizza le caratteristiche del dataset di addestramento e cerca di trovare modelli o relazioni tra gli input e gli output corrispondenti. L'algoritmo cerca di approssimare una funzione che mappa gli input agli output desiderati, in modo da poter fare previsioni corrette per nuovi dati non visti in precedenza.

Laddove il dataset non sia stato adeguatamente curato, non abbia un numero sufficiente di dati o questi non siano adeguatamente rappresentativi, il sistema di IA lavorerà sulla base di dati non completamente corretti e, conseguentemente, avrà dei pregiudizi sulla base dei quali emetterà degli output potenzialmente discriminanti. Al riguardo, esistono già numerosi esempi di pregiudizi relativi ai campi di cui ci stiamo occupando, riguardanti l'origine etnica, il colore della pelle o il genere⁽⁸⁵⁾, ma la casistica non si limita solamente a considerare la razza o il genere, in quanto questi problemi possono verificarsi anche in relazione ad altri elementi, come l'età o le disabilità.

Come già accennato, l'esistenza di tali pregiudizi è legata a molteplici motivi, come l'inadeguatezza dei dataset rappresentativi e la scelta dei metodi e degli algoritmi di apprendimento automatico utilizzati dai *data scientists*⁽⁸⁶⁾. Altre fonti di pregiudizi sono rappresentate dal contesto in cui viene utilizzata l'IA, da pregiudizi storici involontari e da modelli di governance inefficaci. Inoltre, i sistemi sviluppati in contesti ricchi o nei paesi sviluppati potrebbero non essere accurati in contesti con risorse limitate o nei paesi in via di sviluppo.

7. — *Pregiudizi e conseguenze in ambito medico e legale.*

Sulla base di quanto appena specificato, emerge in modo lapalissiano che la costruzione del dataset è fondamentale per avere output corretti. Sistemi

⁽⁸⁵⁾ A. ZAVRŠNIK, *Criminal justice, artificial intelligence systems, and human rights*, ERA Forum 20, 2020, pp. 567-583; W.N. PRICE II, *Medical AI and Contextual Bias*, 33 *Harv. J.L. & Tech.*, University of Michigan Public Law Research Paper No. 632, 2019; O. SCHWARTZ, *In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation*, IEEE spectrum, 2019, in spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation.

⁽⁸⁶⁾ Un data scientist è un professionista specializzato nell'analisi dei dati e nell'applicazione di metodologie e algoritmi statistici per estrarre informazioni significative dai dati. Il ruolo del data scientist implica la raccolta, l'organizzazione, l'elaborazione e l'interpretazione dei dati al fine di identificare modelli, tendenze e insight utili per prendere decisioni informate e guidare il processo decisionale aziendale.

di Intelligenza artificiale mal addestrati potrebbero cagionare gravi danni, soprattutto laddove utilizzate in ambiti quali quello medico e giuridico. Nel settore sanitario, laddove siano coinvolte informazioni legate al fenotipo o al genotipo dei pazienti, un sistema di IA mal addestrato e che presenta pregiudizi potrebbe, ad esempio, suggerire diagnosi errate e rendere conseguentemente inefficaci i trattamenti proposti. Tale inefficacia potrebbe essere assoluta, come avvenuto ad esempio nel caso di *IBM Watson for Oncology*. Questo è un sistema di intelligenza artificiale sviluppato da IBM per supportare i medici nella diagnosi e nel trattamento dei pazienti oncologici. L'obiettivo di *Watson for Oncology* era quello di analizzare una vasta quantità di dati clinici, come pubblicazioni mediche, studi scientifici e report dei pazienti, per fornire raccomandazioni di trattamento personalizzate. Tuttavia, nel corso degli anni il sistema ha ricevuto diverse critiche riguardanti la sua efficacia e accuratezza. Nel 2017, un'indagine di STAT News⁽⁸⁷⁾ ha sollevato dubbi sulla sua capacità di fornire raccomandazioni appropriate, evidenziando casi in cui il sistema avrebbe proposto trattamenti discutibili o in contrasto con le linee guida oncologiche, anche in considerazione del fatto che IBM non aveva prodotto alcuna pubblicazione scientifica relativa alla tecnologia in questione. Ciò ha sollevato preoccupazioni sulla sicurezza e sulla correttezza delle decisioni assunte e delle terapie suggerite, anche conseguentemente a lamentele segnalate da alcuni ospedali che avevano adottato il sistema al di fuori degli Stati Uniti. Questi, infatti, segnalavano la presenza di pregiudizi. Successivamente, diversi ospedali e istituti di ricerca ne hanno sospeso o ridimensionato l'uso e nel 2018 alcuni documenti interni di IBM hanno dimostrato come il sistema spesso forniva consigli errati per il trattamento del cancro e che gli specialisti medici dell'azienda avevano identificato diversi esempi di raccomandazioni di trattamenti non sicuri ed errati⁽⁸⁸⁾. Questi problemi sono probabilmente derivati da errori nella fase di addestramento

⁽⁸⁷⁾ C. ROSS, I. SWETLITZ, *IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close*, StatNews, 2017, in www.statnews.com/2017/09/05/watson-ibm-cancer/.

⁽⁸⁸⁾ C. ROSS, I. SWETLITZ, *IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show*, in StatNews, 2018, www.statnews.com/wp-content/uploads/2018/09/IBMs-Watson-recommended-unsafe-and-incorrect-cancer-treatments-STAT.pdf.

del sistema in quanto invece di utilizzare dati reali di pazienti, l'Intelligenza Artificiale, sembrerebbe essere stata addestrata con pochi dati sintetici, ovvero creati dai medici del Memorial Sloan Kettering Cancer Centre⁽⁸⁹⁾. Questo caso evidenzia le complessità nell'applicazione dell'Intelligenza Artificiale nella pratica clinica e l'importanza di una validazione rigorosa e di una collaborazione stretta tra medici e tecnologi per garantire la sicurezza e l'efficacia di tali strumenti. Contemporaneamente emerge la necessità di una regolamentazione adeguata che possa garantire al contempo che i sistemi immessi in commercio siano robusti e sicuri, che possa tutelare gli utenti ed eventualmente contemplare risarcimenti in caso di danni, ma che sappia anche contemperare questi aspetti con le necessità derivanti dallo sviluppo tecnologico.

Laddove non assoluta, ovvero quando il dataset non è completamente errato, ma presenta dei pregiudizi specifici dovuti alla carenza di informazioni su determinati soggetti o minoranze, l'inefficacia potrebbe essere relativa, avendo ad oggetto solo alcune sottopopolazioni⁽⁹⁰⁾. Questo scenario è possibile in considerazione del fatto che in campo medico l'IA viene addestrata utilizzando i dati disponibili, il che significa che spesso, e soprattutto storicamente, questi riguardano pazienti di origine caucasica⁽⁹¹⁾, generalmente maschi⁽⁹²⁾ e di età adulta⁽⁹³⁾. Tale *bias* non può essere considerato esclusivamente “doloso” in quanto dipendente dal fatto che lo sviluppo della medicina moderna e i relativi dati a disposizione riguardano soprattutto le suddette categorie, a causa della disponibilità derivante da anni di raccolta avvenuta

⁽⁸⁹⁾ S. GERKE, T. MINSSEN, G. COHEN, *Ethical and legal challenges of artificial intelligence-driven healthcare*, cit., pp. 295-336.

⁽⁹⁰⁾ S. GERKE, T. MINSSEN, G. COHEN, *Ethical and legal challenges of artificial intelligence-driven healthcare*, cit., pp. 295-336.

⁽⁹¹⁾ S. COHEN-MEKELBURG, S. BERRY, R.W. STIDHAM, J. ZHU, A.K. WALJEE, *Clinical applications of artificial intelligence and machine learning-based methods in inflammatory bowel disease*, in *J. Gastroenterol Hepatol*, 2021, 36(2), pp. 279-285.

⁽⁹²⁾ S. SHRESTHA, S. DAS, *Exploring gender biases in ML and AI academic research through systematic literature review*, in *Frontiers in Artificial Intelligence*, Volume 5, 2022.

⁽⁹³⁾ J. STYPIŃSKA, A. FRANKE, *AI revolution in healthcare and medicine and the (re-)emergence of inequalities and disadvantages for ageing population*, in *Frontiers in Sociology*, Volume 7, 2023.

nei paesi a più alta industrializzazione, che coinvolgono conseguentemente quel tipo di soggetti. Tale punto di partenza però non può giustificare, oggi, l'utilizzo di dataset non bilanciati. In questo caso, infatti, l'output derivante risulterebbe meno accurato o addirittura inesatto per altre popolazioni per le quali i dati di addestramento sono stati insufficienti, nonostante oggi siano comunque disponibili.

Questo tipo di pregiudizi possono essere evitati utilizzando un numero maggiore di dati e con una variabilità adeguata laddove possibile. Nel caso in cui questo non sia attuabile, a causa di una carenza di dati relativi a specifiche minoranze, dovrebbe essere obbligatorio specificare per quali popolazioni l'algoritmo può essere correttamente utilizzato, per quali potrebbe fornire output non completamente corretti e per quali soggetti non dovrebbe essere utilizzato. Tuttavia, anche in questo caso, il problema principale è rappresentato dagli algoritmi “*black-box*”, in cui la mancanza di trasparenza rende impossibile rilevare i pregiudizi.

Proprio in relazione a questa tipologia di inefficacia esistono dei casi che ci permettono di rilevare una comunanza di problematiche, e di possibili soluzioni, tra i due settori considerati, quello medico e quello giuridico.

Partendo da quest'ultimo settore, un caso di inefficacia relativa delle previsioni dell'IA causata dall'esistenza di pregiudizi riguarda uno dei sistemi più utilizzati negli Stati Uniti, COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*). Trattasi di un software sviluppato da una società privata, utilizzato per stimare il rischio di recidiva degli imputati in alcune giurisdizioni statunitensi. L'algoritmo si basa su una serie di fattori, che possono includere precedenti penali e informazioni demografiche, e sui dati personali dell'imputato per calcolare un punteggio dal quale trarre un livello di rischio utilizzato per desumere la possibilità di recidiva e conseguentemente fornire un output consigliando il giudice circa la decisione da prendere. Questo sistema è stato oggetto di alcuni studi⁽⁹⁴⁾ che, analizzando

⁽⁹⁴⁾ Si veda ad esempio J. ANGWIN, J. LARSON, S. MATTU, L. KIRCHNER, *Machine Bias*, in www.propublica.org; J. DRESSEL, H. FARID, *The accuracy, fairness, and limits of predicting recidivism*, in *Science Advances*, 2018, fasc. 4, p. 20. È opportuno specificare che Equivant (precedentemente nota come Northpointe), una società specializzata nella fornitura di soluzioni software

gli esiti delle decisioni basate sull'utilizzo del software, hanno concluso che le previsioni fatte da questo sistema sono inaffidabili in quanto profondamente compromesse da pregiudizi razziali.

Secondo quanto esposto nel report *Machine bias*, emerge una propensione maggiore a considerare erroneamente gli imputati afroamericani come ad alto rischio di recidiva rispetto agli imputati caucasici, mentre questi ultimi erano più inclini a essere erroneamente identificati come a basso rischio rispetto ai primi. Per giungere a questi risultati sono stati analizzati oltre 10.000 casi della Contea di Broward, in Florida, confrontando i tassi di recidiva previsti con quelli effettivamente verificatisi nel corso di un periodo di due anni ed è quindi emerso che il sistema era in grado di prevedere correttamente la possibilità di recidiva di un imputato nel 61% dei casi. La percentuale di successo scendeva al 20% considerando i soggetti relativi a quella che è stata definita come “recidiva violenta”⁽⁹⁵⁾. Tra gli output del programma, gli imputati di colore sono stati “etichettati”⁽⁹⁶⁾ come ad alto rischio di recidiva più di quanto effettivamente fossero. L'analisi ha rilevato che i soggetti afroamericani che i due anni successivi non hanno compiuto altri reati avevano quasi il doppio delle probabilità di essere erroneamente classificati come ad alto rischio rispetto ai loro omologhi caucasici (45% contro 23%). A parti invertite, gli imputati caucasici venivano invece “etichettati” come meno rischiosi di quanto fossero, in quanto coloro che hanno commesso reati entro i successivi due anni sono stati erroneamente “etichettati” come

per il sistema di giustizia penale e creatrice del sistema COMPAS, ha, in risposta, pubblicato uno studio sostenendo l'erroneità del report *Machine bias*. Cfr. A. FLORES, K. BECHTEL, C. LOWENKAMP, *False Positives, False Negatives, and False Analyses: A Rejoinder to “Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks”*, in *Federal Probation Journal*, September 2016.

⁽⁹⁵⁾ Per recidiva gli estensori del report hanno considerato la definizione data dalla società produttrice di COMPAS, ovvero *a finger-printable arrest involving a charge and a filing for any uniform crime reporting (UCR) code*, mentre per recidiva violenta è stato utilizzata la definizione di crimine violento data dall'FBI, che include omicidio volontario, omicidio colposo, rapina e aggressione.

⁽⁹⁶⁾ Si utilizza il termine “etichettati” in quanto il sistema compie una procedura di *labeling*, o etichettamento, nel classificare i soggetti.

a basso rischio quasi il doppio rispetto alle loro controparti afroamericane (48% contro 28%).

L'analisi ha inoltre dimostrato che, anche considerando i precedenti penali, l'età e il genere, gli imputati afroamericani avevano il 45% in più di probabilità di ricevere punteggi di rischio più alti rispetto a quelli caucasici, e il doppio delle probabilità di questi ultimi di essere erroneamente classificati come ad alto rischio di recidiva violenta. Riguardo gli imputati caucasici, questi avevano il 63% di probabilità in più di essere erroneamente classificati come a basso rischio di recidiva violenta, rispetto ai soggetti afroamericani, i quali avevano il 77% in più di probabilità di ricevere punteggi di rischio più alti.

Appare dunque chiaro come in presenza di questo tipo di differenze, la sussistenza di un pregiudizio razziale, che nel caso di specie si può ritenere derivante anche dalla sovrarappresentazione della popolazione afroamericana nei dataset utilizzati, anche in conseguenza della sovrarappresentazione della stessa popolazione nelle carceri del sistema giudiziario degli Stati Uniti d'America, impatti negativamente sull'intero sistema giudiziario. Questo avviene poiché attraverso l'apprendimento automatico un algoritmo cercherà di massimizzare la precisione delle previsioni per i dati di addestramento. Se un gruppo specifico di individui compare più frequentemente degli altri nei dati di addestramento, il programma si ottimizzerà per quegli individui perché ciò aumenta la precisione complessiva. Gli informatici valutano gli algoritmi utilizzando insiemi di dati di "test", ma di solito questi sono sottoinsiemi casuali dell'insieme di addestramento originale e quindi è probabile che contengano gli stessi pregiudizi. Gli algoritmi difettosi in tal modo possono amplificare i pregiudizi attraverso cicli di feedback⁽⁹⁷⁾. In tal modo si viene

⁽⁹⁷⁾ Ad esempio, nel caso dei sistemi addestrati statisticamente, come avviene per *Google Translate*, si può considerare il problema del genere del pronome. Questo sistema di traduzione di default utilizza il pronome maschile, usando un modello determinato dal rapporto tra i pronomi maschili e i pronomi femminili nei corpora in inglese, che è di 2:1. Oltre a questo dato, sbilanciato di partenza, ogni volta che un programma di traduzione fa riferimento a "*he said*" (lui ha detto), aumenta la frequenza relativa del pronome maschile sul web, incidendo sull'output.

a creare un postulato che tende ad autoavverarsi, in quanto “etichettando” una specifica minoranza e causandone la sovrarappresentazione, il sistema continuerà a ritenere provato il fatto che la suddetta minoranza è ad alto rischio di recidiva, poiché il sistema imparerà dai dati e dalle statistiche che considerano anche le conseguenze dei suggerimenti dati dal sistema stesso.

A conferma del fatto che il problema in oggetto non è relativo solo ad un caso specifico, sono stati rilevati problemi simili in alcuni sistemi di IA utilizzati per determinare l'entità della pena da comminare in alcuni stati⁽⁹⁸⁾, e il tutto è aggravato dal fatto che i soggetti che utilizzano sistemi decisionali automatizzati non hanno accesso all'algoritmo e/o al codice sorgente⁽⁹⁹⁾, creando conseguentemente un problema dal punto di vista della trasparenza di tali sistemi.

Stabilito in tal modo l'impatto sul settore giuridico dell'uso di sistemi di IA portatori di pregiudizi, è interessante valutare se tale problema sussiste anche in altri settori, ed in particolare in quello medico.

Per comprendere come la situazione si stia evolvendo in campo sanitario, è stato possibile far riferimento a una serie di articoli scientifici che hanno affrontato la questione, dato che tale aspetto, a differenza che nel campo giuridico, è già conosciuto ed è stato oggetto di studi approfonditi. Prenderemo ad esempio alcuni studi che hanno dimostrato come l'utilizzo di dataset non bilanciati dal punto di vista del genere impatti sull'efficacia della diagnosi. Questo viene amplificato dal fatto che spesso i dataset utilizzati sono sempre gli stessi, quelli che mettono a disposizione un numero adeguato di informazioni.

A questo proposito, uno studio⁽¹⁰⁰⁾ ha chiarito che le reti neurali usate per la classificazione delle immagini vengono spesso addestrate su ImageNet,

⁽⁹⁸⁾ Alcuni di questi sistemi sono utilizzati, ad esempio, in Arizona, Colorado, Delaware, Kentucky, Louisiana, Oklahoma, Virginia, Washington e Wisconsin.

⁽⁹⁹⁾ M. ŠKOP, M. MERÉNYI, T. TURASHVILI, K. IZDEBSKI, D. KEREKEŠ, V. ILIC, *alGOV-rithms, State of play – Report on Algorithms Usage in Government-Citizens Relations in Czechia, Georgia, Hungary, Poland, Serbia and Slovakia*, Varsavia, 2019.

⁽¹⁰⁰⁾ J. ZOU, L. SCHIEBINGER, *AI can be sexist and racist – it's time to make it fair*, in *Nature*, 2018, 559(7714), pp. 324-326.

che dispone di oltre 14 milioni di immagini etichettate. In tali dataset, frequentemente, alcuni gruppi sono sovrarappresentati e altri sono sottorappresentati. Più del 45% dei dati di ImageNet, proviene dagli Stati Uniti⁽¹⁰¹⁾, che ospitano solo il 4% della popolazione mondiale. Al contrario, Cina e India contribuiscono complessivamente solo al 3% dei dati, nonostante questi paesi rappresentino il 36% della popolazione mondiale. In ambito medico questi problemi risultano ulteriormente amplificati a causa della difficoltà e del costo di produzione e categorizzazione delle immagini mediche. Nel 2017 dei ricercatori che hanno utilizzato il deep learning per identificare il cancro della pelle dalle fotografie hanno addestrato il loro modello su un insieme di dati di 129.450 immagini, delle quali circa il 60% è stato raccolto da Google Immagini⁽¹⁰²⁾. Tuttavia, meno del 5% di queste immagini raffigurano persone dalla pelle scura e, inoltre, l'algoritmo non è stato testato su tali soggetti. Anche in questo caso il possibile pregiudizio tende a creare meccanismi ricorsivi, in quanto sono gli stessi algoritmi che seguiranno lo stesso ciclo che abbiamo identificato in precedenza in ambito giuridico (vedi *infra*).

Tornando agli studi, il primo⁽¹⁰³⁾, pubblicato su BMJ Health & Care Informatics, ha ricreato quattro modelli di intelligenza artificiale documentati in precedenti ricerche noti per avere una percentuale di successo superiore al 70% nell'individuazione delle malattie del fegato dai risultati degli esami del sangue. Conseguentemente sono state analizzate le prestazioni dell'algoritmo in base al genere ed è emerso che la percentuale di diagnosi errate in caso di pazienti di genere femminile era pari al 44% rispetto al solo 23% per gli uomini.

⁽¹⁰¹⁾ S. SHANKAR, Y. HALPERN, E. BRECK, J. ATWOOD, J. WILSON, D. SCULLEY, *No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World*, Workshop on Machine Learning for the Developing World, 2017.

⁽¹⁰²⁾ A. ESTEVA, B. KUPREL, R. NOVOA et al., *Dermatologist-level classification of skin cancer with deep neural networks*, in *Nature*, 2017, 542, pp. 115-118.

⁽¹⁰³⁾ I. STRAW, H. WU, *Investigating for bias in healthcare algorithms: a sex-stratified analysis of supervised machine learning models in liver disease prediction*, BMJ Health Care Inform, 2022, 29(1):e100457.

Anche il secondo studio⁽¹⁰⁴⁾ affronta il problema della poca attenzione dedicata al modo in cui vengono raccolti i database e su come ciò possa influenzare le prestazioni dei sistemi di intelligenza artificiale, mettendo in luce l'importanza dell'equilibrio di genere nei dataset di imaging medico utilizzati per addestrare i sistemi di intelligenza artificiale per la diagnosi assistita dal computer. Basandosi su tre architetture di reti neurali profonde e due dataset noti di immagini a raggi X disponibili e pubbliche utilizzati per diagnosticare varie malattie toraciche in diverse condizioni di squilibrio di genere è stata riscontrato una costante diminuzione delle prestazioni per i generi sottorappresentati quando non viene soddisfatto un equilibrio minimo.

8. — *Conclusioni.*

Questi esempi confermano la tesi per la quale la preparazione e l'utilizzo di dataset sbilanciati comportano un'inefficienza, assoluta o relativa, nei confronti dei soggetti a sfavore dei quali avviene lo sbilanciamento. Inoltre, questo non avviene solo in relazione a uno specifico settore o ad una determinata minoranza, infatti, da un lato accade tanto in ambito medico quanto in quello giuridico e dall'altro coinvolge sia aspetti di genere che razziali.

In questo modo, sebbene le conseguenze siano differenti, afferenti al campo medico o giuridico, in realtà in tutti i casi considerati viene potenzialmente lesa un diritto del soggetto appartenente al gruppo o alla minoranza, sia esso il diritto alla salute o alla libertà personale.

Questa problematica solleva, di conseguenza, una questione che appare essere di primaria importanza e che necessita di essere adeguatamente valutata dai soggetti responsabili della regolamentazione e dell'approvazione dei suddetti sistemi, sia che si occupino di diagnosi assistita dal computer, che di supporto alle decisioni giudiziarie.

Se da un lato, quello medico, esistono però già delle regolamentazioni,

⁽¹⁰⁴⁾ A.J. LARRAZABAL, N. NIETO, V. PETERSON, D.H. MILONE, E. FERRANTE, *Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis*, Proceedings of the National Academy of Sciences, 2020, 117(23), pp. 12592-12594.

anche se dovrebbero includere esplicite raccomandazioni circa l'equilibrio dei dataset sia dal punto di vista del genere che della diversità, dal lato giuridico ci si trova di fronte ad una situazione di assenza di regolamentazione, e perlopiù tali aspetti sono affrontati internamente da linee guida etiche delle imprese che sviluppano e producono i sistemi di IA.

Sarebbe quindi interessante, in considerazione della sussistenza di problematiche comuni, implementare in entrambi i settori delle regole comuni in materia di bilanciamento dei dataset⁽¹⁰⁵⁾, che dovrebbero valere orizzontalmente non solo per questi settori ma per tutti quelli in cui vengono utilizzati sistemi di Intelligenza Artificiale ad alto rischio. Per quello che riguarda il solo settore giuridico, invece, potrebbe essere interessante valutare l'adozione di una specifica regolamentazione calibrata sulla base di quella esistente in ambito medico per i dispositivi medici implementanti l'IA. In tal modo si potrebbe creare un sistema coerente con quanto previsto dalla discutenda normativa unionale e in grado di considerare le specificità e le necessità del settore nonché di essere al passo con lo sviluppo che il progresso dell'IA avrà nei prossimi anni.

⁽¹⁰⁵⁾ Da questo punto di vista *l'Artificial Intelligence Act*, che rappresenta il livello di regolamentazione più avanzato al mondo richiede un'adeguata robustezza dei dataset ma non parla mai di bilanciamento dei dati contenuti.