RIVISTA ITALIANA DI INFORMATICA E DIRITTO
PERIODICO INTERNAZIONALE DEL CNR-IGSG

ISSN 2704-7318 • n. 2/2024 • DOI 10.32091/RIID0175 • articolo sottoposto a peer review • pubblicato in anteprima il 6 nov. 2024
licenza Creative Commons Attribuzione - Non commerciale - Condividi allo stesso modo (CC BY NC SA) 4.0 Internazionale

Tonino Cantelmi – Luisa Lodevole

Nuove prassi e norme per un sistema sanitario "smart": ricadute dell'*AI Act* sulla relazione di cura

L'articolo si propone di esplorare lo stato dell'arte dell'utilizzo dei sistemi di IA nel settore sanitario, in particolare sotto forma di software per il processo decisionale clinico e di dispositivi medici. Gli autori assumono una duplice prospettiva: da un lato, quella del giurista che rileva le implicazioni pratiche delle applicazioni dell'IA nella cura e nella promozione della salute umana e raccomanda l'adozione di una normativa o valuta quella adottata, in questo caso il recente regolamento emanato dall'Unione europea. Dall'altro lato, viene assunto il punto di vista della scienza medica che, unito alla riflessione bioetica, evidenzia gli effetti di ritorno sulla visione della vita dell'uomo postmoderno.

Intelligenza artificiale – Sistema sanitario – Coscienza – Scientismo – Fiducia

New practices and standards for a 'smart' healthcare system: impact of the AI Act on the care relationship

The paper aims to explore the state of the art of the use of AI systems in the healthcare sector, particularly in the form of software for clinical decision-making and medical devices. The authors take a twofold perspective: on the one hand, that of the jurist who notes the practical implications of AI applications in the treatment and promotion of human health and recommends the adoption of a regulation or evaluates the one adopted, in this case, the recent Regulation issued by the European Union. On the other hand, the point of view of medical science is taken, which, combined with bioethical reflection, highlights the feedback effects on post-modern man's view of life.

Artificial intelligence – Healthcare – Consciousness – Scientism – Trust

T. Cantelmi è professore associato all'Istituto di Psicologia, Università Gregoriana – Roma e membro del Comitato Nazionale per la Bioetica, Gruppo di lavoro "Tecnologie emergenti"

L. Lodevole è ricercatrice presso il Dipartimento di Giurisprudenza, Università degli Studi di Roma Tor Vergata

L'articolo è stato concepito e sviluppato in modo condiviso dagli autori. Mentre i paragrafi 1 e 10 sono riferibili ad entrambi, Tonino Cantelmi è autore dei paragrafi nn. 2, 3, 4 e 5; Luisa Lodevole è autore dei paragrafi nn. 6, 7, 8 e 9

Sommario: 1. Introduzione: l'intelligenza artificiale: tra diritto e medicina. – 2. Circuiti elettronici e circuiti cerebrali. – 3. Neuroplasticità e *Intelligenza meccanica*. – 4. *Menti, cervelli* … e coscienze. – 5. Il riduzionismo scientista. – 6. Un modello normativo per i sistemi di IA. – 7. L'attività dei sistemi di IA sotto la lente del diritto: definizioni e classificazioni. – 8. Responsabilità per l'utilizzo dei sistemi di IA in ambito sanitario. – *8.1. L'imputazione soggettiva: le figure coinvolte nella catena di valore dell'IA.* – *8.2. La ricostruzione oggettiva: tra obblighi di trasparenza e* explainability. – 9. IA nel sistema sanitario: da ausilio ad ausiliario. – 10. *AI in healthcare*: un problema di fiducia.

1. Introduzione. L'intelligenza artificiale tra diritto e medicina

Dagli sviluppi della ricerca sull'Intelligenza artificiale[1] (d'ora in poi IA) non sembra impossibile che l'IA in futuro sostituisca il medico. Anzi, già oggi in alcune aree del mondo i pazienti preferiscono l'iscrizione in sistemi di assistenza sanitaria privati[2], digitalizzati e telematizzati, piuttosto che nei servizi sanitari offerti dall'amministrazione pubblica, in particolare il classico medico di medicina generale "in carne ed ossa".

Le applicazioni dei sistemi di IA alle attività di cura e promozione della salute[3] umana si sono moltiplicate negli ultimi anni, così come sono sempre più numerosi gli studi sulle implicazioni deontologiche[4] e bioetiche[5] di questa prassi.

Gli utilizzi delle tecnologie quali ausilio del procedimento clinico, sia nel momento della formulazione della diagnosi[6] che della prescrizione della terapia, oppure come strumento operativo nella chirurgia robotica[7], sollevano questioni di carattere deontologico e biogiuridico[8], nonché riflessioni antropologiche[9] e bioetiche[10].

Una prima questione riguarda la natura e finalità dell'atto medico e le sue leggi intrinseche. Una riflessione sulla specificità dell'arte medica permette di valutare in quali segmenti dell'attività

1. Per una trattazione del tema si veda Amato Mangiameli 2023; Amato Mangiameli 2022; Amato Mangiameli 2022-a. Cfr. anche Finocchiaro 2022; Franceschelli 2023; Gabrielli–Ruffolo 2019; Sartor 2022.
2. Cfr. l'app di servizi di medicina generale già notevolmente diffusa nel Regno Unito "Over 100,000 people already have GP at Hand as their NHS GP, with a 96% satisfaction score", in Babylon GP at hand.
3. Cfr. Tozzi 2023; nonché Alhashmi–Hashem–Al-Qudah 2024; Iqbal–Cortés Jaimes–Makineni et al. 2023; Khera–Butte–Berkwits et al. 2023; Parekh–Shaikh–Simran et al. 2023; Xu–Yang–Arikawa–Bai 2023.
4. Bali–Garg–Bali 2019; Barracca–Ledda–Mancosu et al. 2017; He–Baxter–Xu et al. 2019; Mishra–Savulescu–Giubilini 2021; Musacchio–Guaita–Ozzello et al. 2018.
5. Daverio–Macioce 2023; Morana–Balduzzi–Morganti 2022; Palazzani 2021. Cfr. anche Ferioli 2019.
6. Consiglio Superiore di Sanità 2021.
7. Consiglio Superiore di Sanità 2023.
8. Cfr. nell'ambito del Consiglio d'Europa: Council of Europe 1997; Council of Europe 2020; Mittelstadt 2022. Nel contesto delle Nazioni Unite: UNESCO 2021, UNESCO 2017, UNESCO 2005. Per l'Organizzazione Mondiale della Sanità: World Health Organization 2021.
9. Kampowski 2010.
10. Cfr. il parere espresso da Comitato Nazionale per la Bioetica 2020. Per un inquadramento sistematico Maier–Shibles 2011.

sanitaria l'utilizzo di sistemi di IA sia possibile e auspicabile e in quali non lo sia. Oltre a ciò, consentirebbe di individuare quali fasi eventualmente non possano essere delegate ad un sistema automatizzato, a pena di tradire la natura specifica della professionalità medica.

Un'altra questione emerge a causa del tipo di tecnologia: i sistemi di IA, software o robot[11], sollevano, rispetto a tutte le altre tecnologie informatiche e telematiche, dilemmi nuovi, derivanti dalle caratteristiche proprie di un agente intelligente, che non solo è in grado di pensare umanamente ma anche di agire razionalmente[12] e di continuare a modificare e riprogrammare se stesso attraverso l'apprendimento automatico.

L'ipotesi di macchine pensanti[13] ha suscitato da lungo tempo linee di ricerca volte ad esplorare il funzionamento della mente umana[14] e la riproducibilità del pensiero[15]: sul punto, sebbene le scienze neurologiche e cognitive abbiano analizzato le circonvoluzioni della massa cerebrale e scoperto alcune reazioni elettrochimiche dei suoi circuiti, allo stato attuale l'investigazione è appena agli inizi e non sembra che sia stata trovata la via d'uscita dal labirinto.

Un ulteriore interrogativo sorge a causa dell'intrusione di un ulteriore "agente" nell'interazione tra medico e paziente[16], il sistema di IA, che, però, non è un mero *medium*, ma neanche un soggetto[17]. La mediazione di questo "terzo" in una relazione che è finora stata sempre "a due", potrebbe portare ad esiti, in una certa misura, imprevedibili: eppure, poiché l'arte medica si è in larga parte consolidata in buone pratiche e linee guida, ci si chiede come questo elemento terzo possa essere incluso nei protocolli e nelle prassi cliniche o chirurgiche. D'altro canto, poiché quelle linee guida costituiscono il parametro di valutazione della diligenza[18] del professionista – ben diversa dall'ordinaria diligenza del comune cittadino – il giurista deve risolvere due fondamentali *quaestiones*: quali siano i criteri di imputazione delle attività svolte dall'agente intelligente[19]; se siano applicabili, e in che modo, i regimi previsti dalle legislazioni nazionali in tema di responsabilità professionale in campo sanitario anche quando intervenga la mediazione dell'IA.

Il diritto, con le sue categorie rigide e fredde, non riesce a modulare il suo intervento in base alla diversa intensità di cooperazione dei differenti sistemi di IA: quindi risulta particolarmente difficile inquadrare l'IA con i modelli classici del diritto civile.

2. Circuiti elettronici e circuiti cerebrali

La ricerca sull'IA ha avuto origine dal tentativo di riprodurre il pensiero in macchine. Gli studiosi a partire dal XX secolo hanno mosso i primi passi su questa strada con diversi approcci.

Da un lato cercarono di osservare e descrivere cosa accade nel cervello umano quando pensa, con tecniche di indagine, c.d. *neuroimaging*, sempre più sofisticate, che ad oggi non hanno ancora esaurito tutte le questioni aperte.

Dall'altro lato, come prospettato da Turing, i ricercatori del settore hanno tentato di riprodurre le attività del cervello tramite ingranaggi o

11. Per una classificazione nelle due categorie di SIAP ovvero "Sistemi di IA Programmi" e SIAR "Sistemi di IA Robot", cfr. Taddei Elmi 2020, in particolare p. 848.
12. Cfr. Russell–Norvig 2021, pp. 4-8.
13. Turing 1950.
14. Cfr. lo studio di Searle 1984.
15. Il binomio intelligenza artificiale fu utilizzato per la prima volta nel 1956, anno della *Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* primo evento ufficialmente volto a riunire ricercatori su questa linea di ricerca, in particolare John McCarthy, Marvin Minsky, Claude Shannon e Nathaniel Rochester, cfr. McCarthy–Minsky–Rochester–Shannon 2006. Gli studi di McCulloch e Pitts sul neurone artificiale e quelli di Turing sulle macchine pensanti sono da considerare prodromici rispetto all'incontro del 1956. Per le notizie storiche v. Russell–Norvig 2021, p. 19 ss.
16. Helzel 2021.
17. Per una discussione sul punto cfr. Sartor 2002 e Sartor 2022, nonché Taddei Elmi 2020.
18. Campiglio 2023; Grasso 2023; nonché Ruffolo 2021; Ruffolo 2021-a.
19. Lucarelli Tonini 2023; Salito 2022; Salvatore 2023.

meccanismi che potessero "lavorare" come il cervello, ovvero elaborare le informazioni in ingresso, sulla base di una tavola di istruzioni, allo stesso modo in cui le elabora la mente umana, fornendo gli stessi prodotti in uscita. La sfida per gli studiosi consisteva pertanto nell'arrivare a codificare una tavola di istruzioni tale che, inserita in un cervellone elettronico, gli permettesse di funzionare e quindi di risolvere problemi allo stesso modo in cui li avrebbe risolti un uomo. Trovata la tavola di istruzioni, si sarebbe potuto dire di aver scoperto come funziona il cervello.

Si ricorda che il test di Turing era costruito come una prova *a contrario*: come il matematico esplicitamente affermava[20], egli stesso non era in grado di esporre argomenti sufficienti per fondare l'ipotesi che le macchine possono pensare, pertanto, la sua arringa a sostegno dei suoi argomenti era articolata come confutazione delle principali controdeduzioni a lui rivolte dalle controparti.

Di conseguenza, come evidenziato da Searle[21], anche nel caso in cui la macchina riuscisse ad ingannare l'interrogante[22] – il quale non fosse in grado di distinguere quale risposta al suo quesito provenisse da un uomo e quale da una macchina – la prova di Turing non dimostrava affatto che la macchina potesse pensare, ma che potesse simulare le attività del cervello umano, «manipolando simboli formali»[23].

In questo senso le osservazioni di Searle, secondo il quale il computer può benissimo superare la prova di Turing senza "comprendere nulla", possono essere condivise. Al fine di equiparare la macchina alla mente, infatti, non sarebbe sufficiente che la macchina potesse riprodurre gli stati cognitivi della mente, poiché ciò che caratterizza gli stati mentali della mente umana è l'intenzionalità, che dipende dai procedimenti causali e non da quelli formali:

> Il problema, col simulatore del cervello, è che simula le cose sbagliate del cervello. Finché simula solo la struttura formale della sequenza delle aggregazioni di neuroni e della sinapsi, non avrà simulato ciò che importa nel cervello: precisamente le sue proprietà causali, la sua abilità a produrre stati intenzionali[24].

I sostenitori dell'IA forte, secondo John Searle, non si avvedono di riproporre nelle loro teorie un "dualismo mascherato"[25] in base al quale *res cogitans* e *res extensa*, di cartesiana memoria, sarebbero, seppure non empiricamente, almeno teoricamente separabili, in modo tale che sia possibile riprodurre il funzionamento della mente indipendentemente dal sostrato causale del cervello.

D'altro canto, secondo l'autore, gli stessi sostenitori dell'IA forte peccherebbero di un certo comportamentismo, e questo vizio di fondo indurrebbe gli stessi a ritenere di aver riprodotto e reso osservabile la causa degli stati cognitivi, che per l'uomo è nel cervello, per il solo fatto di aver riprodotto, mediante manipolazione di simboli formali, gli stati cognitivi o processi mentali nel calcolatore elettronico:

> Poiché i computer, programmati appropriatamente, possono avere modelli input-output simili a quelli degli esseri umani, siamo tentati di postulare che essi abbiano stati mentali simili agli stati mentali umani. Ma una volta che vediamo che è possibile, sia concettualmente che empiricamente, che un sistema abbia capacità umane in qualche campo senza avere alcuna intenzionalità, dovremmo poter superare questo impulso [26]

Se si ripercorre il ragionamento dei comportamentisti in effetti si può riscontrare che il sillogismo è viziato, in quanto si fonda su una premessa che postula un'equivalenza inesistente: «la mente sta al cervello come il software sta

20. Turing 1950, pp. 148-149: «Il lettore si sarà già accorto che non ho alcuna argomentazione molto convincente di carattere positivo per sostenere il mio punto di vista. Se l'avessi avuta non mi sarei certo dedicato con tanta cura a indicare gli errori dei punti di vista opposti al mio».

21. Searle 1984, p. 55.

22. Turing 1950, pp. 121-122.

23. Searle 1984, p. 58.

24. *Ivi*, p. 60.

25. *Ivi*, p. 71.

26. *Ivi*, p. 69.

all'hardware e così possiamo capire la mente senza fare neurofisiologia»[27].

3. Neuroplasticità e *Intelligenza meccanica*

Ciò che si propone questo studio è mettere a disposizione di altre discipline, come il diritto o la bioetica, i risultati degli studi sul cervello, al fine di demistificare alcune proposizioni indimostrate e del tutto infondate.

In particolare, sembra che oggi sia dato per assiomatico l'assunto secondo il quale riprodurre in calcolatori elettronici alcune attività e procedure, alcune manifestazioni del cervello, implichi e giustifichi due ulteriori affermazioni: 1) che possiamo dire ormai che il cervello sia conoscibile, come si conosce un oggetto empirico; 2) che il cervello si riduca a quelle attività o funzioni.

Se tali affermazioni fossero vere, allora i discorsi di Searle sull'intenzionalità cadrebbero, in quanto nulla di differente e specifico avrebbero gli stati cognitivi umani rispetto agli stati cognitivi della macchina.

Al contrario, gli studi condotti negli ultimi decenni sul cervello che interagisce con le tecnologie digitali hanno evidenziato alcune caratteristiche peculiari del cervello, in particolare una, che è nota come neuroplasticità[28].

Questo significa che il nostro cervello non solo è in grado di apprendere ed elaborare informazioni, ma esso cambia continuamente a causa delle esperienze e acquisizioni, sia cognitive, sia neuro-sensoriali.

Lo stesso vissuto, i progetti, le percezioni e i comportamenti, ancor più se ripetuti, conformano e modellano i circuiti cerebrali.

Le macchine non hanno esperienze allo stesso modo in cui gli uomini le hanno, poiché gli uomini conoscono e sperimentano attraverso la corporeità che è parte integrante e inscindibile, dell'essere persona[29].

4. *Menti, cervelli* … e coscienze

Se si prende sul serio la qualità neuroplastica[30] dei circuiti cerebrali, si comprende come anche la conoscenza non sia un'attività solo mentale[31] ma risieda in molti luoghi corporei diversi, seppure a volte si localizzi in transito a partire da alcune terminazioni neuro-sensoriali (quando tocchiamo il fuoco e subiamo una bruciatura; quando vediamo un lampo o ascoltiamo un tuono, etc.).

Alcuni studi in psicologia hanno approfondito ciò che la disciplina chiama "schema corporeo"[32] per sottolineare che l'attività conoscitiva non è solo intellettuale, ma passa per il nostro corpo e le nostre azioni. Queste ricerche hanno evidenziato che spesso non troviamo «dove finiamo noi e dove inizia il resto del mondo», ovvero la nostra coscienza, perché «l'abbiamo cercata nel posto sbagliato»[33].

> I nostri corpi e le nostre menti sono attive. Cambiando la forma della nostra attività possiamo cambiare la nostra stessa forma, il nostro corpo e la nostra mente … Oggi pensiamo all'economia come a qualcosa di globalizzato, alle società per azioni come a qualcosa di internazionalizzato e alle reti informatiche come qualcosa di diffuso. Noi stessi siamo per natura dinamicamente distribuiti, in grado di oltrepassare i confini e situati in un ambiente[34].

Riprendendo Searle, il fatto che possiamo riprodurre determinate operazioni mentali o stati

27. *Ivi*, p. 59.
28. Pascual Leone–Amedi–Fregni–Merabet 2005.
29. Spaemann 2013; cfr. sul punto gli studi di antropologia biblica di Frevel 2009, pp. 258-260: «… Il cuore può essere sinonimo di volontà, progetto, senso, sentimento. Oltre alle funzioni di tipo emotivo e volitivo, si legano al cuore pure quelle noetiche. L'uomo veterotestamentario non pensa con la testa ma con il cuore … Il cuore è la sede dell'intelletto, della ragione, della coscienza, della facoltà cognitiva, della percezione, della memoria, della capacità di apprendere, riflettere e giudicare, della ponderazione argomentata».
30. Cfr. Song–Millidge–Salvatori et al. 2024.
31. Cfr. Gandolfini 2013; nonché Palazzani–Zannotti 2013.
32. Il riferimento è a Noë 2010, p. 81.
33. *Ivi*, p. 72.
34. *Ivi*, p. 71 ss.

cognitivi in un procedimento computazionale non significa che quel modello computazionale svolga quelle operazioni allo stesso modo del cervello umano[35].

Questo non solo è evidente a livello esperienziale[36], ma l'inferenza secondo la quale se un procedimento computazionale mi fornisce lo stesso risultato di un'operazione mentale, allora tutte le operazioni mentali non sono altro che procedimenti computazionali, è teoricamente errata, per diversi ordini di motivi.

In primo luogo, non tutte le operazioni mentali sono riproducibili da una macchina, ovvero non potrei trasferire la modalità umana di pensare nella macchina, ma solo far sì che questa simuli le sequenze logiche e i procedimenti causali onde ottenere lo stesso prodotto.

L'errato sillogismo, che è alla base del pensiero contestato da Searle, sarebbe formulabile all'incirca come segue:

premessa A – se un'attività è riproducibile da una macchina è (espressione della natura) meccanica,

+ premessa B – se gli stati mentali umani sono riproducibili da una macchina,

= conclusione – dunque il cervello è una macchina.

In realtà gli errori sono presenti in più di un passaggio: nell'articolo determinativo "gli" che induce a ritenere che "tutti" gli stati mentali siano riproducibili da una macchina, piuttosto che solo "alcuni"; nel salto logico che attribuisce al cervello una natura, sulla base di un tipo di attività che svolge. Il cervello può esprimersi in una molteplicità di attività diverse, delle quali solo una minima parte è formulabile in modelli computazionali; allo stesso modo se un animale, come un pappagallo o una scimmia, può riprodurre un comportamento umano, non vuol dire che l'imitazione dell'azione implichi l'uguaglianza degli agenti.

In secondo luogo, il fatto che la macchina possa eseguire alcune attività umane, non solo logiche, ma anche pratiche, non è sufficiente ad inferire che la macchina possa essere equiparata all'uomo e, viceversa, l'uomo a una macchina.

5. Il riduzionismo scientista

La prospettiva meccanicista e scientista, sopra rappresentata, risulta infondata. Non soltanto l'attività cerebrale non è conoscibile come quella di un oggetto allo studio delle sole scienze empiriche, perché le nostre menti sono "estese"[37] e sono integrate nel nostro "schema corporeo". Ma neanche l'attività mentale stessa è riducibile a quella meccanica.

Dalle sopra esposte considerazioni emerge con vigore la questione antropologica nel rapporto tra tecnologie e essere umano: l'utilizzo e lo sviluppo dell'IA non costituisce tanto un problema bioetico in quanto strumento per molteplici attività, tra le quali le attività sanitarie, ma nella misura in cui pone le basi per un "cambiamento d'epoca"[38] dell'umano.

Secondo l'approccio scientista[39] posso conoscere solo ciò che posso osservare e misurare, di conseguenza la realtà, anche l'uomo, è ridotto a ciò che di lui posso osservare e misurare. La deriva di questa teoria consiste nel negare l'esistenza di tutto ciò che non è misurabile empiricamente, con presupposto ideologico e atteggiamento antiscientifico: ciò che non posso misurare, non esiste.

Il *bias* del riduzionismo scientifico consiste nel dedurre dallo studio del cervello e dalla riproducibilità di alcune operazioni logiche in calcolatori elettronici, che il cervello non sia altro che una macchina.

35. Searle 1984, p. 69: «Il fatto che il programmatore e l'interprete dell'output usano i simboli come sostituti di oggetti nel mondo è totalmente al di fuori degli obiettivi del computer. Il computer, per ripeterci, ha una sintassi ma non una semantica. … Il punto non è che esso manca di qualche informazione di secondo ordine per l'interpretazione dei suoi simboli di primo ordine, ma piuttosto che i suoi simboli di primo ordine non hanno interpretazioni almeno per quanto riguarda il computer. Tutto ciò che il computer ha, sono simboli».
36. Cfr. Gallagher–Zahavi 2009.
37. Cfr. Noë 2010.
38. Francesco 2022.
39. Cowburn 2013.

Il successivo passaggio riduzionista, indotto dalla *Cognitive Science*[40], sarebbe che il funzionamento del cervello esaurisce il pensiero e la riflessione prudenziale: anche questo risulta essere un assunto apodittico, poiché le scienze neurologiche e psicologiche dimostrano che l'uomo è molto più delle sue capacità cognitive e che il suo cervello produce un pensiero personale, che la macchina mai potrà riprodurre. Chiamiamo coscienza[41] ciò che si sottrae alla misurazione e all'osservazione empirica.

L'errore delle scienze cognitive, evidenziato da Searle, è che esse intendono studiare il funzionamento della mente, sulla base del presupposto che il pensiero sia pur sempre un'attività astratta e separata rispetto al cervello, ovvero che «la mente sia un tipo di computer»[42]. Nonostante quelle teorie ritengano che «non sia sufficiente simulare le operazioni della mente per spiegare l'intelligenza umana»[43], alla fine riducono pur sempre il pensiero ad operazioni su codici simbolici.

Anche se le macchine diventassero capaci di apprendere e «staccarsi dal comportamento prevedibile implicato nel calcolo», ovvero diventassero intelligenti nel senso in cui lo intende Turing, ciò non cambierebbe il discorso di fondo, e cioè che è ingiustificato inferire dall'attività logico-meccanica che la causa di essa sia, o possa diventare, un'intelligenza del tipo di quella umana: rimarrà pur sempre un'*intelligenza meccanica*.

6. Un modello normativo per i sistemi di IA

Il possibile utilizzo dell'IA nell'ambito sanitario solleva alcuni quesiti giuridici ulteriori, rispetto a quelli che sorgono in qualsiasi altro settore, a causa delle specificità e complessità dell'atto medico, che può essere definito come condotta – guidata da conoscenze tecnico-scientifiche "fisico-chimico-biologiche" e innestata su "dinamiche relazionali"[44] – finalizzata alla cura, alla promozione della vita e della salute o del completo benessere psico-fisico del paziente.

Assumendo la prospettiva sopra descritta, al fine di comprendere le ricadute giuridiche sulla relazione medico-paziente delle applicazioni di IA, è necessario chiedersi in quali fasi dell'attività di cura si collochino le operazioni degli agenti intelligenti artificiali e se questi siano autonomi centri di imputazione di atti *lato sensu* medici; se integrino una componente modale o strumentale – un ausilio – dell'atto imputabile al professionista; o ancora se possano essere ricondotti alla figura dell'ausiliario, a causa della parziale autonomia dei processi che da essi si originano.

L'impianto normativo dell'*AI Act*, Regolamento UE 2024/1689[45], non affronta questi quesiti in modo diretto, poiché non inquadra il sistema di IA dal punto di vista del diritto sostanziale, bensì ne regola l'utilizzo dal punto di vista procedurale.

Infatti, sebbene sia stata sostenuta l'ipotesi dell'attribuzione di una soggettività giuridica al sistema di IA, come persona giuridica dotata di un patrimonio o *e-person*, questa soluzione non è stata avallata né dalla dottrina prevalente né tantomeno dal legislatore europeo. L'analisi del giurista deve pertanto limitarsi a ricavare i collegamenti e comporre un quadro d'insieme tra la recentissima regolamentazione sull'IA e la disciplina vigente a livello euro-unitario e nazionale sulla responsabilità del produttore e sull'attività sanitaria. In questo modo l'adozione di una normativa sull'IA, che chiarisca obblighi e responsabilità dei diversi soggetti coinvolti, dovrebbe conseguire l'effetto di

40. Lepore–Pylyshyn 1999.
41. Cfr. Bayne–Cleeremans–Wilken 2009.
42. Pylyshyn 1999, p. 12.
43. *Ibidem*, la traduzione è dell'autore.
44. Bellino 2015.
45. Il testo del Regolamento (UE) 2024/1689 è stato approvato dal Parlamento europeo in data 13 marzo 2024 e dal Consiglio dell'Ue in data 21 maggio; è stato pubblicato nella *Gazzetta Ufficiale dell'Unione europea* in data 12 luglio 2024 e, alla data del 16 luglio 2024. Ormai esauriti i passaggi dell'iter procedurale, sarà applicabile dal 2 agosto 2026, ma l'art. 6 comma 1 a partire dal 2 agosto 2027, cfr. Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale).

ingenerare sicurezza e fiducia negli utilizzatori e, allo stesso tempo, incentivare il progresso tecnologico e lo sviluppo del mercato.

Dal punto di vista giuridico è necessario comprendere in che modo e misura i movimenti dei robot o gli output di un software possano concorrere ad integrare la condotta del professionista o operatore sanitario: se costituiscano un *medium* – un'estensione del medico o un suo potenziamento – oppure a pieno titolo un agente al pari di un ausiliario, al fine di poter rispondere alle domande sopra esposte.

L'*AI Act* enuncia esplicitamente le finalità della regolamentazione adottata, ovvero quelle di «migliorare il funzionamento del mercato interno e promuovere la diffusione di un'intelligenza artificiale (IA) antropocentrica e affidabile»[46] garantendo la tutela dei diritti fondamentali, dello Stato di diritto, della democrazia in conformità ai valori dell'Unione.

L'attenzione specifica di giuristi e legislatori è stata richiamata durante gli ultimi decenni, nei quali si sono sviluppate queste sofisticate tecnologie, dal fatto che esse possono incorporare meccanismi decisionali automatizzati, che mettono a rischio interessi e beni giuridici primari, protetti dagli ordinamenti nazionali e internazionali: la vita e la salute umana; la sicurezza di tutti gli operatori sanitari, del paziente e dell'umanità intera; la deontologia della professione medico-sanitaria; l'innovazione tecnologica e gli investimenti industriali; gli interessi di mercato.

Per comporre queste posizioni giuridiche, diritti e interessi, in un delicato equilibrio, sono stati necessari diversi anni, durante i quali si sono svolti i lavori preparatori[47] e sono stati interpellati organi consultivi[48] e gruppi di esperti[49], al fine di pervenire alla stesura di una prima proposta nel 2021[50] ed alla redazione dell'articolato attuale.

Già il testo del 2021 – che è stato in gran parte confermato dal provvedimento di recente approvato, seppur con alcuni emendamenti di rilievo – presceglieva l'impianto normativo dei sistemi di gestione dei rischi o *New Legislative Framework*[51].

Questo tipo di struttura regolatoria richiede a coloro che intraprendono un'attività connotata da rischi di vario genere, ineliminabili in radice, di costruire un modello di organizzazione e stabilire determinate procedure che, pur contemplando l'ipotesi di verificazione del rischio[52], cercano di

46. Art. 1, comma 1.
47. Si fa riferimento in ambito Ue alle Comunicazioni e proposte succedutesi negli ultimi dieci anni, tra le quali principalmente: Comunicazione della Commissione al Parlamento europeo, al Consiglio europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, *Piano coordinato sull'intelligenza artificiale,* 7 dicembre 2018, COM(2018) 795; Comunicazione della Commissione al Parlamento europeo, al Consiglio europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, *Creare fiducia nell'intelligenza artificiale antropocentrica,* 8 aprile 2019, COM(2019) 168; Commissione europea, *Libro bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia,* 19 febbraio 2020, COM(2020) 65.
48. European Group on Ethics in Science and New Technologies 2018.
49. High-Level Expert Group on Artificial Intelligence 2019, 2019-a, 2020. Si veda anche lo studio Joint Research Centre 2020.
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, 21 aprile 2021, COM(2021) 206.
51. Lo schema normativo del *New Legislative Framework* è descritto in dettaglio nel Regolamento (CE) n. 765/2008 del Parlamento europeo e del Consiglio del 9 luglio 2008 che pone norme in materia di accreditamento e vigilanza del mercato per quanto riguarda la commercializzazione dei prodotti e che abroga il regolamento (CEE) n. 339/93.
52. Nei sistemi di gestione dei rischi gli eventi che integrano la "non conformità" o discostamente dallo standard, sono i più vari, a seconda del campo al quale si applica l'apparato di prevenzione, ad es.: nel campo industriale, malfunzionamento o errato funzionamento di macchinario con esito dannoso nei confronti di cose o persone; in attività commerciali, l'illecito commesso da un dipendente d'azienda; nelle aziende di tutti i tipi, l'infortunio del lavoratore; negli istituti di formazione e istruzione, la lesione, volontaria o involontaria, dello studente nei locali o pertinenze della scuola, ecc.

limitarne la probabilità di ricorrenza e ridurne al minimo l'impatto.

L'assunzione di una percentuale di alea – di scostamento dallo standard o esito atteso –, come *condicio sine qua non* dell'esercizio dell'attività *de qua*, comporta un ribaltamento di prospettiva da parte del legislatore: la disciplina non può, a meno di vietare del tutto la prassi, descrivere astrattamente la verificazione del rischio come ipotesi "patologica" di per sé indicativa di una colpa, bensì può assumere come presupposto che, chiunque "agisca" o "subisca" l'attività pericolosa, abbia accettato quel rischio e le connesse conseguenze giuridiche in termini di imputazione e regime di responsabilità.

La normativa pertanto stabilisce modelli procedurali e requisiti formali ai quali i produttori – ma anche nelle altre fasi i fornitori[53], assemblatori, gestori, sviluppatori e utilizzatori – devono attenersi, per ottenere la conseguente valutazione di conformità del sistema di IA e l'autorizzazione all'immissione in circolazione.

Similmente alla circolazione dei veicoli a motore, per i quali è previsto l'obbligo assicurativo[54] a carico del proprietario, così anche per le attività, che si servono di sistemi di IA, l'autorizzazione all'immissione in circolazione del robot o software non garantisce l'assenza di rischio, bensì assolve a una precisa *ratio legis*, quella di prevedere ed attuare un complesso di misure per tenere l'alea sotto controllo e ridurla al minimo.

Dall'esame degli articoli da 8 a 15 e da 16 a 27[55], si evince la struttura del modello, ovvero una serie di controlli preventivi, precauzioni e sanzioni, oltre che l'assoggettamento ad una sorveglianza continua. All'articolo 9 rubricato "Sistemi di gestione dei rischi", precisamente al comma 2 sono descritte in sequenza le quattro operazioni principali in cui si sostanzia la disciplina: identificazione e analisi dei rischi; valutazione della loro probabilità di evenienza; monitoraggio successivo e verifiche periodiche; trattamento dei rischi al fine dell'obiettivo del miglioramento continuo.

Quello descritto è, in forma molto semplificata, lo schema normativo dei sistemi di gestione dei rischi delle organizzazioni (aziendali), paradigma che si sta affermando in molteplici settori non solo industriali, tramite normative cosiddette "tecniche" (peraltro anche in settori nient'affatto tecnici).

7. L'attività dei sistemi di IA sotto la lente del diritto: definizioni e classificazioni

La definizione di sistema di IA contenuta all'articolo 3, n. 1 deve ormai essere considerata normativa: questo è «un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali».

La grande partizione introdotta dalla normativa è data dalla distinzione tra le diverse tipologie di rischio indotto dai sistemi di IA: alcuni di essi realizzano attività vietate, descritte all'articolo 5, perché implicano un livello o qualità di rischio inaccettabile; altre pratiche, di cui agli articoli 6-7, sono ritenute a rischio accettabile ma elevato, per l'effetto i sistemi che le dispiegano sono definiti "ad alto rischio" e assoggettati a verifiche e requisiti rigorosi; tutti gli altri sono definiti in via residuale a rischio basso o minimo.

Il divieto[56] di produzione e di utilizzo, previsto dall'articolo 5, vige per quei sistemi di IA considerati, di per sé o per l'uso che se ne può fare, particolarmente pericolosi, come chiarisce il

53. Cfr. D'Adda 2022.

54. Cfr. la Risoluzione del Parlamento europeo del 2017 sulla robotica, che ipotizza al n. 58 l'applicazione ai robot del regime di assicurazione obbligatoria vigente per i veicoli a motore: «ritiene che, come avviene nel caso dell'assicurazione dei veicoli a motore, tale regime assicurativo potrebbe essere integrato da un fondo per garantire la possibilità di risarcire i danni in caso di assenza di copertura assicurativa», Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).

55. Rispettivamente Sezioni 2 e 3 del Capo III.

56. Art. 5 "Pratiche di IA vietate" in particolare riferimento ai sistemi di IA che attribuiscono un punteggio sociale in base ai comportamenti individuali, cfr. Cerrina Feroni 2023.

Considerando 28[57], in riferimento a quei sistemi per i quali il rischio per la salute, la sicurezza e i diritti fondamentali è considerato inaccettabile.

In questa sede sembra rilevante menzionare tra i sistemi vietati quelli manipolatori (lett. a) del comma 1 – in grado di influire sulla volontà e le scelte dei singoli – e classificatori (lett. c) del comma 1 delle persone, in grado di ingenerare effetti distorsivi a livello sociale, poiché regolano, ad esempio, con criteri discriminatori l'accesso o la distribuzione di beni e di servizi, in special modo se adottati da pubbliche amministrazioni per la gestione di servizi pubblici essenziali. Oltre a ciò, sono inclusi nell'elenco dei sistemi vietati quelli che consentono «l'identificazione biometrica "in tempo reale" in spazi accessibili al pubblico» (lett. h), comma 1 senza alcuna informazione a coloro che vi sono coinvolti e senza che sussista una delle situazioni previste a giustificazione di eccezioni al divieto.

L'articolo 6 invece qualifica come ad alto rischio il sistema che «presenti un rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche, anche nel senso di influenzare materialmente il risultato del processo decisionale» [58].

Per i sistemi di IA ad alto rischio il Regolamento prevede determinati requisiti – rispetto ai quali i produttori, fornitori e utilizzatori devono essere *compliant*, come l'adozione di un sistema di gestione dei rischi (cfr. art. 9) – affinché possano essere immessi sul mercato o in servizio[59].

Nonostante il comma 1 dell'articolo 6 statuisca la regola generale che assoggetta i sistemi "ad alto rischio" ad una serie rigorosa di controlli, monitoraggi e verifiche ed il comma 2 qualifichi di default "ad alto rischio" tutti i sistemi appartenenti ad uno dei settori di cui all'allegato III[60], tuttavia il comma terzo dello stesso articolo introduce una deroga di non poco conto. Tale norma stabilisce che alcuni sistemi, pur da considerare "ad alto rischio" perché rientranti in una delle categorie di cui all'elenco dell'allegato III – tra i quali, ciò che qui interessa, i dispositivi medici – sono eccezionalmente esclusi dalla qualificazione e sottratti alla relativa disciplina, qualora sussista una delle condizioni espressamente specificate dalle lettere da *a*) a *d*).

Trattandosi di norma speciale che deroga a quella generale, la lettura del testo, integrata dal superiore comma 2, produce la seguente affermazione: alcuni sistemi di IA, che dovrebbero essere considerati ad alto rischio, perché incidenti su diritti fondamentali o interessi di rilievo pubblicistico, non sono considerati tali, purché sussista una o più delle seguenti condizioni: «a) il sistema di IA è destinato a eseguire un compito procedurale limitato; b) […] a migliorare il risultato di un'attività umana precedentemente completata; c) […] a rilevare schemi decisionali o deviazioni da schemi decisionali precedenti e non è finalizzato a sostituire o influenzare la valutazione umana […]; d) […] a eseguire un compito preparatorio per una valutazione pertinente ai fini dei casi d'uso elencati nell'allegato III»[61].

57. Considerando 28: «L'IA presenta, accanto a molti utilizzi benefici, la possibilità di essere utilizzata impropriamente e di fornire strumenti nuovi e potenti per pratiche di manipolazione, sfruttamento e controllo sociale. Tali pratiche sono particolarmente dannose e abusive e dovrebbero essere vietate poiché sono contrarie ai valori dell'Unione relativi al rispetto della dignità umana, alla libertà, all'uguaglianza, alla democrazia e allo Stato di diritto e ai diritti fondamentali sanciti dalla Carta, compresi il diritto alla non discriminazione, alla protezione dei dati e alla vita privata e i diritti dei minori».
58. Art. 6. *Regole di classificazione per i sistemi di IA ad alto rischio*.
59. Per la differenza tra "immissione sul mercato", "messa a disposizione sul mercato" e "messa in servizio" confronta art. 3 "Definizioni", rispettivamente nn. 9, 10 e 11.
60. Si ricordano in breve le otto categorie ad alto rischio *ex* Allegato III: «Biometria…, Infrastrutture critiche…, Istruzione e formazione professionale…, Occupazione, gestione dei lavoratori e accesso al lavoro autonomo…, Accesso a servizi privati essenziali e a prestazioni e servizi pubblici essenziali…, Attività di contrasto…, Migrazione, asilo e gestione del controllo delle frontiere…, Amministrazione della giustizia e processi democratici …».
61. Art. 6, comma 3 che riprende il Considerando 53.

Questa norma solleva rilievi critici sia quanto alla formulazione, alquanto tortuosa, sia quanto al contenuto, poiché le deroghe introdotte aprono all'interpretazione spazi troppo ampi. Non è chiaro cosa voglia dire che un sistema di IA, seppure appartenente ad una delle categorie di cui all'allegato III, non sia da considerare ad alto rischio, perché debba «eseguire un compito procedurale limitato»[62] e chi decida su questa qualificazione. La stessa osservazione vale per gli altri tre casi di deroga, da b) a d), poiché queste disposizioni implicano valutazioni ampiamente discrezionali attribuendo un minor livello di pericolosità a sistemi che incidono, *ab origine* a causa del settore in cui operano, su diritti fondamentali, sulla salute o sulla sicurezza dei cittadini.

8. Responsabilità per l'utilizzo dei sistemi di IA in ambito sanitario

Sebbene appaiano di grande portata i possibili benefici dei sistemi di IA applicati al procedimento clinico, in termini di velocità della diagnosi, personalizzazione e precisione delle terapie, nondimeno un malfunzionamento o errore, c.d. *bias*, del sistema può determinare conseguenze anche gravi, o letali, a danno della salute o dell'integrità fisica delle persone.

La questione del danno prodotto da sistema di IA per produzione di output errato o mancata produzione dell'output, riguarda il tema della responsabilità.

8.1. L'imputazione soggettiva: le figure coinvolte nella catena di valore dell'IA

Il legislatore europeo mediante l'adozione del *NLF* affronta le fondamentali questioni giuridiche insorte a seguito dello sviluppo delle applicazioni di IA – principalmente quelle dell'imputazione delle attività nelle diverse fasi, dalla programmazione alla messa in servizio, e delle conseguenti responsabilità – in modo indiretto, disponendo a carico di coloro, che entrano nella catena del valore dell'IA[63] – soprattutto per i sistemi ad alto rischio – obblighi e adempimenti, a seconda del ruolo che assumono nell'assemblaggio del prodotto "finito".

Già è stato accennato che il modello del sistema di gestione dei rischi – applicato a molteplici enti quali aziende, industrie, società commerciali e finanziarie, operanti in diversi settori – in ossequio alla logica dei sistemi di gestione, non ha lo scopo di stabilire con precisione il limite tra il lecito e l'illecito, ma di imporre all'organizzazione (aziendale, ecc.) di collocare ogni sequenza della propria attività in una complessa procedura e di mantenerla sotto un costante monitoraggio, allo scopo di rilevare in quale sezione si verifichi eventualmente la difformità o scostamento dallo standard, al fine di adottare misure correttive e realizzare il miglioramento continuo.

Nei diversi comparti di attività d'impresa una certa tipologia o percentuale di rischi – fotografati all'inizio della costruzione del "modello" organizzativo nel documento chiamato "analisi dei rischi" o *risk assessment* – è insita e prevedibile fin dall'origine. La cosiddetta "non conformità" o allontanamento dall'iter procedurale atteso rappresenta una fattispecie astratta che di volta in volta può essere integrata da atti o fatti molto diversi tra loro: azioni od omissioni di ausiliari (compresa la commissione di reati); malfunzionamenti di attrezzature o impianti, oppure errori di utilizzo, che provocano danni a terzi.

Trasferendo questa logica alla produzione dei sistemi di IA, considerata la difficoltà di ipotizzare preventivamente tutte le diverse correlazioni[64] e implicazioni tra soggetti che compartecipano alla fabbricazione di un sistema siffatto, è comprensibile che il Regolamento non affronti in modo diretto il principale problema in punto di diritto, che è quello dell'imputazione e conseguente responsabilità dei danni provocati dai sistemi di IA, ma rinvii in forma implicita alla normativa europea e nazionale in merito, ovvero attenda che la Commissione si esprima con un proprio orientamento.

Lo schema normativo dell'*AI Act* fornisce purtuttavia il quadro dei soggetti obbligati ad adottare misure preventive e correttive dei rischi, di talché la mancata osservanza delle formalità, comporta per ciò stesso un comportamento colpevole ed un correlato addebito.

62. Art. 6 comma 3.

63. Questa espressione compare nel testo del Regolamento per la prima volta al Considerando 9.

64. Per un'analisi approfondita delle figure che partecipano alla produzione dei sistemi di IA e dei criteri di imputazione delle responsabilità in caso di verificazione di danni v. D'Adda 2022.

Ad un primo esame del testo si può notare come l'attenzione del legislatore sia sbilanciata verso le figure che intervengono nella parte finale della sequenza, nella fase della produzione e immissione in circolazione del prodotto: ben si comprendono le motivazioni che hanno indotto ad assumere tale prospettiva, ovvero quelle di favorire negli utilizzatori un convincimento sull'affidabilità di questi sistemi e di promuoverne la circolazione sul mercato, rendendo agevole per l'utente l'azionabilità di un'eventuale domanda di ristoro del danno.

Sebbene questa impostazione fosse stata suggerita già da una proposta di Regolamento sulla responsabilità per funzionamento dei sistemi di IA del 2020[65], tuttavia essa non ha avuto seguito, perché effettivamente troppo gravosa e penalizzante per gli operatori finali. Si pensi alle ricadute sulla relazione di cura di una siffatta scelta legislativa. Ciò provocherebbe in materia di responsabilità risarcitoria in ambito sanitario una legittimazione processuale passiva generalizzata delle strutture sanitarie e dei medici (salvo azione di rivalsa), con conseguente incentivo ad un'attitudine "difensiva" dei professionisti e generazione di una diffusa sfiducia nella categoria.

Da un'analisi accurata emerge invece che altri soggetti possono fornire un apporto, anche rilevante, nella messa a punto di una componente del sistema, se non dell'intero prodotto: si tratta degli ideatori dell'algoritmo, programmatori, operatori di *front-end* o di *back-end*[66], assemblatore finale (che può essere produttore del robot e coincidere con il *deployer* o meno), fornitori[67], *deployer*[68], importatori[69], distributori[70] e gli utilizzatori finali.

Per i robot autonomi di intelligenza artificiale emerge il fondamentale apporto di chi partecipa alla ideazione del software, «procede con la codificazione – intesa come sviluppo dell'algoritmo così ideato in un codice interpretabile dalla macchina – ed infine involge un'attività di collaudo/testing»[71]. La medesima impresa che realizza il software potrebbe poi installarlo su un hardware, oppure immetterlo sul mercato, affinché possa essere sviluppato da un altro produttore di robot oppure dallo stesso utilizzatore finale.

Sul fornitore di un sistema ad alto rischio grava l'obbligo di adozione di un sistema di gestione dei rischi, con tutti gli adempimenti[72] connessi, descritti negli articoli 9 e seguenti, come l'obbligo di trasparenza (v. *infra*), affinché siano conoscibili e conosciute dagli utenti le informazioni sulle caratteristiche tecniche dei prodotti e sui rischi potenziali (art. 11).

65. Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale (2020/2014(INL)).

66. Cfr. D'Adda 2022, p. 816: «operatore, inteso come il soggetto che esercita un certo grado di controllo sul sistema di AI e che beneficia del suo funzionamento (si tratta dell'operatore di *front-end*) ovvero il soggetto che definisce le caratteristiche della tecnologia e fornisce i dati od il servizio di supporto al robot (l'operatore di *back-end*)».

67. Sui fornitori cfr. Considerando 79: «È opportuno che una specifica persona fisica o giuridica, definita come il fornitore, si assuma la responsabilità dell'immissione sul mercato o della messa in servizio di un sistema di IA ad alto rischio, a prescindere dal fatto che tale persona fisica o giuridica sia la persona che ha progettato o sviluppato il sistema».

68. Art. 3 n. 4

69. Sugli obblighi degli importatori, cfr. artt. 2, 20 e 23 e il Considerando 83.

70. Art. 24.

71. D'Adda 2022, p. 826.

72. I sistemi ad alto rischio devono corrispondere ai requisiti di accuratezza, robustezza e cibersicurezza, come già indicato dal Considerando 74: «Le prestazioni dei sistemi di IA ad alto rischio dovrebbero essere coerenti durante tutto il loro ciclo di vita e tali sistemi dovrebbero garantire un livello adeguato di accuratezza, robustezza e cibersicurezza, alla luce della loro finalità prevista e conformemente allo stato dell'arte generalmente riconosciuto. La Commissione e le organizzazioni e i portatori di interessi pertinenti sono incoraggiati a tenere in debita considerazione l'attenuazione dei rischi e degli impatti negativi del sistema di IA. Il livello atteso delle metriche di prestazione dovrebbe essere dichiarato nelle istruzioni per l'uso che accompagnano il sistema».

Lo stesso fornitore potrebbe aver internalizzato[73] tutte le varie fasi della produzione, ma anche aver solo assemblato il sistema alla fine: la differente parcellizzazione o accentramento dei processi determina una difficoltà, da risolvere caso per caso, nella imputazione soggettiva dell'attività dell'IA, di particolare rilievo se lesiva di diritti altrui.

Di particolare importanza la figura, introdotta da ultimo nel testo, del *deployer* che

> … dovrebbe essere interpretata come qualsiasi persona fisica o giuridica, compresi un'autorità pubblica, un'agenzia o altro organismo, che utilizza un sistema di IA sotto la sua autorità, salvo nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale. A seconda del tipo di sistema di IA, l'uso del sistema può interessare persone diverse dal *deployer*[74]…

Nonostante alcune precisazioni definitorie sulle diverse figure operanti nella "catena del valore dell'IA", la legislazione di recente approvata prende in considerazione la difficoltà di individuare ruoli e responsabilità poiché

> in determinate situazioni tali operatori [nda: gli importatori] potrebbero agire contemporaneamente in più di un ruolo e dovrebbero pertanto adempiere cumulativamente tutti gli obblighi pertinenti associati a tali ruoli. Ad esempio, un operatore potrebbe agire contemporaneamente come distributore e importatore[75].

Dalla complessità del quadro disegnato, emerge che un malfunzionamento di un sistema di IA potrebbe derivare dall'erroneo utilizzo dell'operatore finale, ma anche da un errore nella programmazione oppure nell'assemblaggio.

8.2. La ricostruzione oggettiva: tra obblighi di trasparenza e *explainability*

Un'altra criticità dei sistemi di IA, in particolare di quella generativa, è la cosiddetta "opacità intrinseca"[76] ovvero l'estrema difficoltà, per non dire l'impossibilità, di ricostruire *ex post* il processo svolto dall'agente intelligente per produrre l'output a partire dai dati in ingresso. Questa caratteristica rende ulteriormente difficoltosa la ricognizione oggettiva delle sequenze procedurali seguite dal sistema e della causa di un eventuale malfunzionamento.

Dal punto di vista tecnico questo problema era stato individuato fin dall'inizio dallo stesso Turing il quale nel suo celeberrimo saggio del 1950 doveva ammettere che la sfida, di cui all'epoca non si poteva intravedere la soluzione, era quella di educare o addestrare un calcolatore digitale allo stadio di "macchina-bambino" attraverso una tavola di istruzioni, ricompense e sanzioni, in modo da fornirgli non solo le regole per eseguire le operazioni da svolgere ma la propensione ad apprendere e formulare nuove regole[77]. Questo processo di addestramento avrebbe dovuto in ogni caso rendere possibile ad ogni stadio intermedio di avere «una chiara immagine mentale dello stato della macchina in ogni momento del calcolo»[78]. Nonostante la teoria fosse chiara, Turing doveva osservare che «una caratteristica importante di una macchina che impara è che il suo insegnante ignorerà spesso in gran parte ciò che di preciso si verifica al suo interno, quantunque possa essere in grado di predire in qualche misura il comportamento del suo allievo»[79].

La strategia di contrasto all'opacità, ampiamente discussa in fase di lavori preparatori dagli organi legislativi europei nelle diverse stesure, è stata tradotta a livello normativo nella formula

73. Questo è il fenomeno dei c.d. "sistemi chiusi", ovvero imprese nelle quali viene ideato, progettato, programmato e assemblato il prodotto completo, cfr. D'Adda 2022, pp. 826-827.
74. Considerando 13.
75. Considerando 83.
76. Cfr. Messina 2023, la quale distingue tra opacità tecnica, giuridica e intrinseca, essendo quest'ultima: «un fenomeno che si manifesta allorquando è la macchina a dettare autonomamente la regola, rendendo di fatto inaccessibile, anche agli stessi programmatori, l'iter che conduce all'esito dell'elaborazione delle informazioni», pp. 279-280.
77. Turing 1950, p. 148 ss.
78. *Ivi*, p. 155.
79. *Ibidem*.

dell'obbligo di trasparenza, gravante sui produttori e programmatori a tutela dei diritti degli utilizzatori e dei terzi, nei limiti strettamente necessari per non scoraggiare dall'altro lato l'innovazione tecnologica e l'invenzione industriale, protette da diritti di privativa.

Già la proposta del 2021 prevedeva all'art. 52 obblighi di trasparenza per tutti i sistemi di IA, non solo quelli ad alto rischio. La normativa da ultimo approvata richiama fin dalle prime battute questo requisito, che già l'*High-Level Expert Group on Artificial Intelligence* nelle sue *Ethics guidelines for trustworthy AI*[80] aveva enunciato come uno dei sette principi cardine della regolamentazione sull'IA: «intervento e sorveglianza umani, robustezza tecnica e sicurezza, vita privata e governance dei dati, trasparenza, diversità, non discriminazione ed equità, benessere sociale e ambientale e responsabilità». L'attuale Considerando 27, rinviando al *Parere del Gruppo di Esperti*, definisce la trasparenza come il fatto

> che i sistemi di IA sono sviluppati e utilizzati in modo da consentire un'adeguata tracciabilità e spiegabilità, rendendo gli esseri umani consapevoli del fatto di comunicare o interagire con un sistema di IA e informando debitamente i *deployer* delle capacità e dei limiti del sistema e le persone interessate dei loro diritti[81]

Il legislatore europeo affronta chiaramente le "preoccupazioni relative all'opacità", prima che nell'articolato in senso stretto, nel Considerando 72[82] ed insiste sulla spiegabilità dei sistemi di IA e sugli obblighi informativi che gravano sui programmatori e sui fornitori a monte dell'immissione in circolazione. Al fine di realizzare l'obiettivo della trasparenza, i sistemi di IA ad alto rischio

> dovrebbero essere progettati in modo da consentire ai *deployer* di comprendere il funzionamento del sistema … e … essere accompagnati da informazioni adeguate sotto forma di istruzioni per l'uso.

Non solo è previsto l'obbligo di corredare il sistema di istruzioni per l'uso, ma sono altresì imposte

> misure di sorveglianza umana, comprese le misure volte a facilitare l'interpretazione degli output del sistema di IA da parte dei *deployer*.[83]

L'obbligo teorico si concretizza dunque in una serie di adempimenti menzionati all'articolo 1, e descritti in dettaglio negli articoli 13 e 50 (oltre che richiamati in modo sparso in altri luoghi del testo). Si ricorda che, mentre l'articolo 13 si riferisce solo ai sistemi di IA ad alto rischio, l'art. 50 si riferisce in generale a "determinati sistemi di IA", espressione che indica anche altri sistemi, a rischio basso o minimo.

L'articolo 13 rubricato "Trasparenza e fornitura di informazioni ai *deployer*", che nell'ultima formulazione sostituisce agli utenti i *deployer*, specifica il contenuto obbligatorio delle istruzioni per l'uso, richiedendo espressamente ai fornitori di includere nel documento alcuni elementi che consentano la spiegabilità e l'interpretabilità dell'output. In questo senso si sottolineano alcuni interventi innovativi rispetto alla proposta di Regolamento del 2021, quali le lettere b) *iv*) e *vii*) dell'art. 13, comma 3 nonché la lettera f):

> Art. 13. … 3. Le istruzioni per l'uso contengono almeno le seguenti informazioni: b) le caratteristiche, capacità e i limiti delle prestazioni del sistema di IA ad alto rischio, tra cui: … *iv*) se del caso, la capacità e caratteristiche tecniche del

80. Cfr. High-Level Expert Group on Artificial Intelligence 2019. Ai sette principi guida per un'IA affidabile fa espresso riferimento il Considerando 27: «Sebbene l'approccio basato sul rischio costituisca la base per un insieme proporzionato ed efficace di regole vincolanti, è importante ricordare gli orientamenti etici per un'IA affidabile del 2019 elaborati dall'AI HLEG indipendente nominato dalla Commissione. In tali orientamenti l'AI HLEG ha elaborato sette principi etici non vincolanti per l'IA che sono intesi a contribuire a garantire che l'IA sia affidabile ed eticamente valida. I sette principi comprendono: intervento e sorveglianza umani, robustezza tecnica e sicurezza, vita privata e governance dei dati, trasparenza, diversità, non discriminazione ed equità, benessere sociale e ambientale e responsabilità. […] Secondo gli orientamenti dell'AI HLEG con "intervento e sorveglianza umani" si intende che *i sistemi di IA sono sviluppati e utilizzati come strumenti al servizio delle persone, nel rispetto della dignità umana e dell'autonomia*» [il corsivo è redazionale].

81. Considerando 27.

82. Nel testo della proposta del 2021 corrisponde al n. 47, seppur con significative modifiche.

83. Considerando 72.

> sistema di IA ad alto rischio connesse alla fornitura di informazioni pertinenti per spiegarne l'output; …*vii*) se del caso, informazioni che consentano ai *deployer* di interpretare l'output del sistema di IA ad alto rischio e di usarlo in modo opportuno; … f) se del caso, una descrizione dei meccanismi inclusi nel sistema di IA ad alto rischio che consente ai *deployer* di raccogliere, conservare e interpretare correttamente i log in conformità dell'articolo 12[84]

Poiché si tratta di requisiti della fase di programmazione, l'obbligo riguarda i fornitori[85], come chiarito dall'art. 8, comma 2, anch'esso introdotto *ex novo* nell'ultima stesura. Nelle istruzioni per l'uso devono essere presenti una serie di altri elementi[86], tra i quali di particolare interesse le misure di sorveglianza umana, di cui al successivo articolo 14, previste alla lettera d), che richiama di nuovo la necessaria interpretabilità dell'output: «comprese le misure tecniche poste in essere per facilitare l'interpretazione degli output dei sistemi di IA ad alto rischio da parte dei *deployer*».

L'articolo 50 invece è rivolto a fornitori e *deployer* di "determinati sistemi di IA", che sono indicati nei successivi paragrafi da 1 a 4, ad esempio quelli «destinati a interagire *direttamente* con le persone fisiche» oppure sistemi di IA, «compresi i sistemi di IA per finalità generali, che generano contenuti audio, immagine, video o testuali sintetici», ed altri.

Poiché non è possibile in questa sede un approfondimento sulla categoria dei "modelli di IA per finalità generali"[87], si sottolinea che essa è stata introdotta dagli emendamenti del giugno 2023, mentre non era presente nella proposta del 2021.

In conclusione, stante l'impossibilità di stabilire *a priori* e in astratto criteri di attribuzione della responsabilità per danni, si dovrà risalire nel caso concreto allo stadio della procedura nel quale si è verificata la difformità e alla figura di volta in volta imputabile soggettivamente.

Considerata da un lato la difficoltà dell'individuazione dei *tortfeasors*, soggetti obbligati ad ottemperare nelle diverse fasi a prescrizioni e adempimenti; dall'altro la caratteristica opacità che rende difficile dal punto di vista oggettivo la ricostruzione dei processi, un regime di responsabilità per danni che facesse ricadere sul danneggiato l'onere di provare la colpa del soggetto imputabile del malfunzionamento (che potrebbe derivare persino dallo stadio della programmazione) – a meno che non sia derivato da macroscopica imperizia dell'utilizzatore finale – e il nesso di causalità tra condotta colpevole e danno, risulterebbe oltremodo gravoso.

Nella disamina[88] dei diversi regimi di responsabilità applicabili, ad esempio, nell'ordinamento civilistico italiano, si potrebbe discutere dei *pro* e *contra* relativi all'adozione del regime di responsabilità extracontrattuale oppure di un'imputazione

84. Dal testo del Regolamento 2024/1689.

85. Nel Considerando 84 è chiarito che «Al fine di garantire la certezza del diritto, è necessario chiarire che, a determinate condizioni specifiche, qualsiasi distributore, importatore, *deployer* o altro terzo dovrebbe essere considerato un fornitore di un sistema di IA ad alto rischio e, pertanto, assumere tutti gli obblighi del caso», e cfr. Considerando 79.

86. Art. 13, comma 3 lettere da a) a f): «a) l'identità e i dati di contatto del fornitore e, ove applicabile, del suo rappresentante autorizzato; … c) eventuali modifiche apportate al sistema di IA ad alto rischio e alle sue prestazioni che sono state predeterminate dal fornitore al momento della valutazione iniziale della conformità …; e) le risorse computazionali e di hardware necessarie, la durata prevista del sistema di IA ad alto rischio e tutte le misure di manutenzione e cura, compresa la relativa frequenza [*omissis*]; f) se del caso, una descrizione dei meccanismi inclusi nel sistema di IA ad alto rischio [*omissis*]».

87. Nel testo definiti dall'art. 3, n. 63: «un modello di IA, anche laddove tale modello di IA sia addestrato con grandi quantità di dati utilizzando l'autosupervisione su larga scala, che sia caratterizzato una generalità significativa e sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere immessi sul mercato».

88. Cfr. Colletti 2021; D'Adda 2022, p. 830 ss.

a titolo di responsabilità oggettiva per i danni derivanti da sistemi di IA.

La scelta del regime di responsabilità extracontrattuale è stata caldeggiata da una proposta di direttiva[89] sulla responsabilità civile extracontrattuale da attività dei sistemi di IA, pur con il correttivo di una presunzione circa il nesso di causalità tra difetto e danno, per non gravare il danneggiato di un onere della prova troppo difficoltoso. L'elezione di una disciplina della responsabilità a titolo di *strict liability* sarebbe invece sostenuta dall'inquadramento dell'attività dei sistemi di IA tra le attività pericolose *ex* art. 2050 c.c. oppure nella fattispecie della responsabilità per danno cagionato da cose in custodia *ex* art. 2051 c.c., o ancora della responsabilità per atti commessi da dipendenti e ausiliari *ex* art. 2049 c.c. Qualora fosse riconosciuto il sistema di IA come prodotto, sarebbe inoltre possibile l'ascrizione della responsabilità al produttore per danno da prodotto difettoso sia ai sensi della normativa nazionale, di cui al Codice del consumo[90], sia ai sensi della normativa europea[91] stabilita dalla Direttiva 85/374/CEE, della quale è in discussione una proposta di revisione[92].

Un'attenta ponderazione dovrebbe condurre a prescegliere un regime di responsabilità agevole per il danneggiato il quale potrebbe agire nei confronti dell'ultimo utilizzatore, di modo che non sia scoraggiato l'utilizzo di questi sistemi, salvo poi la proponibilità dell'azione di regresso del *solvens* nei confronti "del" o "dei" coobbligati, secondo criteri di riparto della responsabilità che tengano conto dei diversi regimi (in deroga all'art. 2055 c.c.), sulla base dei quali rispondono i (plurimi) condebitori, a titolo di colpa o a titolo oggettivo, per non gravare dall'altro lato l'utilizzatore dell'intera responsabilità per il semplice fatto di aver introdotto il rischio.

Inoltre per calibrare in modo equo la responsabilità di ciascuno dei soggetti che hanno contribuito alla catena produttiva sarà necessario «accertare il ruolo delle attività riconducibili a tali soggetti nella causazione del sinistro»; pertanto la quota di ciascuno nell'obbligazione risarcitoria «si fonderà sul grado della colpa associabile alle diverse condotte, sull'intensità dell'apporto causale dispiegato da ciascuno, e sulla dimensione del rischio riconducibile ad ogni *tortfeasor*»[93].

Nel caso di sistemi utilizzabili in ambito sanitario la gravità del rischio è data dall'importanza dei beni in gioco ovvero la vita e la salute: questo giustifica il superamento del criterio di imputazione a titolo di responsabilità contrattuale *ex* art. 1218 c.c. – che ammette la prova liberatoria con onere a carico del danneggiante, applicabile nell'ordinamento italiano alle ipotesi di responsabilità della struttura sanitaria o del medico libero professionista – a favore del regime di responsabilità a titolo di *strict liability*.

9. IA nel sistema sanitario: da ausilio ad ausiliario

Ad oggi sono moltissime le applicazioni di IA nel campo della medicina e promozione della salute, sia in campo clinico[94], sia in campo chirurgico[95], sia mediante l'utilizzo del "gemello digitale"[96], valido modello artificiale di organo o organismo, per testare le reazioni del paziente reale a interventi e trattamenti.

I sistemi di IA software possono essere utilissimi ausili per la diagnosi delle patologie e la

89. 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) COM(2022) 496.

90. D.lgs. 6 settembre 2005, n. 206.

91. Direttiva 85/374/EEC del Consiglio del 25 luglio 1985 relativa al ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati membri in materia di responsabilità per danno da prodotti difettosi, ancora in vigore nella versione aggiornata.

92. Proposta di Direttiva del Parlamento Europeo e del Consiglio sulla responsabilità per danno da prodotti difettosi COM(2022) 495.

93. D'Adda 2022, p. 836.

94. Chary–Boyer–Burns 2021; Piliuk–Tomforde 2023; Teng–Liu–Song et al. 2022.

95. *Ex multis* Wang–Gao–Váncza et al. 2019.

96. Popa–van Hilten–Oosterkamp–Bogaardt 2021.

formulazione delle terapie in modo personalizzato[97], in considerazione dell'enorme mole di dati che il cervello elettronico ha immagazzinato in fase di addestramento e che può confrontare con la fattispecie concreta in un lasso di tempo minimo. Una delle specialità di maggiore esito riguarda il trattamento del diabete[98].

I dispositivi medici, dotati di un componente di sicurezza che è un sistema di IA o che sono essi stessi sistemi di IA, sono già oggetto di regolamentazione da parte dei Regolamenti 2017/745 e 2017/746[99], ricompresi nell'elenco dell'allegato I dell'*AI Act*, sezione A, "Elenco della normativa di armonizzazione" ai nn. 11 e 12.

Nel caso dei dispositivi medici, costituiti anche da un sistema di IA o che sono essi stessi un sistema di IA, può verificarsi un'ipotesi di duplicazione o contraddizione tra norme: quelle della *lex specialis* che prescrive standard di sicurezza e obblighi di valutazione di conformità[100] per i dispositivi medici e quella della *lex generalis* che riguarda i sistemi di IA componenti di sicurezza di un dispositivo medico o che sono essi stessi un dispositivo medico.

Al fine di prevenire una possibile antinomia il legislatore europeo ha statuito:

> La classificazione di un sistema di IA come ad alto rischio a norma del presente regolamento non dovrebbe necessariamente significare che il prodotto il cui componente di sicurezza è il sistema di IA, o il sistema di IA stesso in quanto prodotto, sia considerato "ad alto rischio" in base ai criteri stabiliti nella pertinente normativa di armonizzazione dell'Unione che si applica al prodotto. Ciò vale, in particolare, per i regolamenti (UE) 2017/745 e (UE) 2017/746, in cui è prevista una valutazione della conformità da parte di terzi per i prodotti a medio rischio e ad alto rischio.[101]

Il Considerando 51 è norma di coordinamento poiché risolve il problema della doppia valutazione di conformità, alla quale sono sottoposti determinati apparecchi utilizzati in ambito sanitario, dotati di componenti di IA: la soluzione dettata dalla legislazione rende indipendenti le due valutazioni che tengono conto di tipologie di rischio differenti e pertanto non necessariamente la qualifica ad alto rischio secondo la normativa speciale dei Regolamenti del 2017 (*Medical Devices Regulations* o MDR) implica la stessa qualificazione nei termini dell'*AI Act*.

Questi particolari sistemi di IA sono assoggettati agli stessi requisiti ed obblighi di tutti gli altri, a seconda che siano inquadrati tra quelli "a rischio alto"[102] o medio/basso.

Per i sistemi ad alto rischio i fornitori, e gli altri soggetti corresponsabili della messa a punto del prodotto finale, devono ottemperare a tutto quanto prescritto dal Capo III del Regolamento, senza possibilità di deroga alla norma generale che richiede la valutazione di conformità, di cui all'art. 43.

Appare degna di menzione la disciplina dell'art. 46, comma 7 che esclude dalla deroga di cui all'art. 43, i sistemi di IA di cui all'allegato I, pur ammettendo in generale che un sistema di IA possa essere messo in commercio o in servizio senza aver superato la procedura di valutazione della conformità, per «motivi di sicurezza pubblica o di protezione della vita e della salute delle persone e di protezione dell'ambiente…»[103].

97. Briganti–Le Moine 2020.

98. Deutsch–Ahlqvist–Udler 2022.

99. I due regolamenti sono pienamente applicabili dal maggio 2021: Regolamento (Ue) 2017/745 del Parlamento Europeo e del Consiglio 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; Regolamento (Ue) 2017/746 del Parlamento Europeo e del Consiglio relativo ai dispositivi medico-diagnostici in vitro e che abroga la direttiva 98/79/CE e la decisione 2010/227/UE della Commissione.

100. Cfr. art. 5, comma 1 Regolamento (Ue) 2017/745: «Un dispositivo può essere immesso sul mercato o messo in servizio solo se è conforme al presente regolamento qualora sia debitamente fornito e correttamente installato, oggetto di un'adeguata manutenzione e utilizzato conformemente alla sua destinazione d'uso».

101. Considerando 51.

102. Cfr. Considerando 46.

103. Articolo 46.

È da notare che l'art. 59 del Regolamento 2017/745 e l'articolo 54 del Regolamento 2017/746 già prevedevano una deroga alla valutazione di conformità richiesta in base ai rischi specifici di quei settori di attività. Tuttavia, mentre la proposta di legge del 2021[104] consentiva la deroga alla procedura di valutazione della conformità anche per i dispositivi medici dotati di IA[105], la formulazione attuale non permette, in un'ottica tuzioristica, che le autorità competenti autorizzino, seppur motivatamente, l'immissione sul mercato o la messa in servizio, nel territorio dello Stato membro interessato, di dispositivi specifici per i quali le procedure di valutazione di conformità non siano state espletate, neanche se l'impiego fosse «nell'interesse della salute pubblica o della sicurezza o salute dei pazienti».

In particolari situazioni di urgenza la normativa deve bilanciare interessi giuridicamente rilevanti di carattere opposto: da un lato l'esigenza dei pazienti a trattamenti e protocolli sempre più evoluti ed efficaci nel minor tempo possibile, dall'altro l'irrinunciabile funzione di garanzia dell'ordinamento giuridico che prescrive un sistema di verifica dei requisiti e plurimi controlli – che richiedono tempo – prima dell'immissione in circolazione di nuovi dispositivi, per promuovere la sicurezza, affidabilità e protezione dei diritti fondamentali. Da un lato l'interesse della ricerca scientifica all'avanzamento delle conoscenze e delle tecnologie, dall'altra la prospettiva tuzioristica del sistema sanitario che deve minimizzare i rischi per i pazienti e la collettività e autorizzare nuovi trattamenti o tecnologie solo quando le percentuali di rischio per la vita o la salute degli esseri umani siano di gran lunga inferiori rispetto ai possibili benefici; da un lato l'interesse di mercato dell'industria biomedica, dall'altro l'interesse della collettività ad accedere al minimo costo e in condizioni di parità alle cure ritenute maggiormente efficaci e più sicure allo stato dell'arte.

Altra disposizione specifica riguarda l'obbligo di notifica degli eventi avversi[106], che, in questo caso, riguarda solo quelli gravi.

Quanto agli obblighi di trasparenza, nel settore sanitario si aggiungono alcune osservazioni rispetto a quanto detto per tutti i sistemi di IA.

Per i sistemi di IA utilizzati nel settore sanitario, è auspicabile il massimo dell'accuratezza per ottenere i migliori risultati possibili dall'applicazione, però questo potrebbe compromettere la trasparenza[107]. Tuttavia, secondo alcuni studiosi[108] non necessariamente una maggiore spiegabilità riduce l'accuratezza, come a prima vista ritenuto. Infatti, una maggiore interpretabilità o trasparenza potrebbe dar luogo ad una maggiore accuratezza[109], perché una migliore comprensione dell'output dei sistemi di IA da parte dei medici, ad esempio della diagnosi prospettata o terapia proposta, potrebbe rendere più efficace la prassi medica ed anche più accettabile dal paziente.

Non sempre la velocità, disponibilità, precisione, ma anche freddezza, automatismo e spersonalizzazione sono preferiti al rapporto interpersonale di fiducia con il medico.

In ultima analisi tutti i sistemi ad alto rischio sono assoggettati ad una disciplina che impone requisiti e procedure per la certificazione di qualità e sicurezza: eppure non è possibile impedire in assoluto che sistemi di IA, che sono caratterizzati dall'apprendimento automatico, in particolar modo se di IA generativa, evolvano in modo da oltrepassare i paletti imposti dalle norme e di conseguenza ledere determinati diritti fondamentali. Questa possibilità in ambito sanitario potrebbe avere un impatto ancora più negativo.

Se l'intelligenza artificiale modifica le procedure e le sequenze di operazioni in corso d'opera,

104. Articolo 59, comma 1, Regolamento 2017/745.

105. Cfr. art. 47, comma 6 della Proposta di Legge sull'Intelligenza artificiale (2021)206.

106. Articolo 73, comma 10: «Per i sistemi di IA ad alto rischio che sono componenti di sicurezza di dispositivi, o sono essi stessi dispositivi, disciplinati dai regolamenti (UE) 2017/745 e (UE) 2017/746, la notifica è limitata agli incidenti gravi di cui all'articolo 3, punto 44, lettera c), del presente regolamento ed è trasmessa all'autorità nazionale competente scelta a tal fine dagli Stati membri in cui si è verificato l'incidente».

107. D'Elia 2023.

108. Cfr. Hatherley–Sparrow–Howard 2023.

109. In questo senso D'Elia 2023.

divorando non solo dati, ma codici e istruzioni su come formulare i codici, non si può fare affidamento su una certificazione di qualità che attesta la regolarità del sistema ad un certo momento, ma non può assicurarne la sicurezza nel prossimo futuro (c.d. *state of art defence*). In effetti il sistema di gestione dei rischi, come noto, può prevedere solo ciò che accade nel maggior numero di casi o ciò che potrebbe presuntivamente accadere, ma non può tenere conto di esiti di un processo di cui non si conosce la procedura.

Con questa consapevolezza anche lo stesso dettato dell'articolo 9[110] potrebbe essere ritenuto oscuro: non appare chiaro come la normativa possa demandare ad un organismo di certificazione una valutazione su qualcosa che effettivamente quest'ultimo non può conoscere ma solo ipotizzare. La stessa espressione "rischi individuati" mira a restringere il campo della gestione del rischio al solo novero di quelli conosciuti, perché potrebbero esservene altri "non individuati", ovvero non conosciuti o conoscibili.

La logica di sistema di gestione del rischio non può soddisfare tutte le esigenze di garanzia e protezione dei diritti, almeno in riferimento ai sistemi di IA caratterizzati dal *deep learning*[111] e IA generativa, poiché essi sono qualificati proprio dalla non-conoscibilità del possibile futuro sviluppo (fenomeno della c.d. *black-box*). Come la procedura civile insegna, non c'è affidabilità senza spiegabilità, perché una decisione è impugnabile solo sulla base delle motivazioni e non del solo dispositivo. Se la decisione di IA sulla diagnosi, ad esempio, diventa una "sentenza senza motivi", risulta difficile anche solo commentarla perché non se ne comprendono le argomentazioni fondative né il percorso logico che giustifica il disposto finale.

Il legislatore europeo, adottando il *New Legislative Framework*, formula una serie di norme procedurali, quasi un "codice di procedura algoritmica", al quale però sfuggono le sequenze degli agenti intelligenti artificiali: nel settore sanitario un difetto di *explainability* e *interpretability* potrebbe portare a conseguenze rilevanti, come provocare sfiducia.

10. AI in healthcare: un problema di fiducia

Quanto fin qui detto conduce la riflessione ad alcune considerazioni finali.

L'utilizzo delle applicazioni di IA in medicina può apportare benefici ed opportunità enormi in termini di supporto alle attività diagnostiche, terapeutiche, chirurgiche e sanitarie in generale. Ai vantaggi si affiancano alcune questioni bioetiche[112] e biogiuridiche analoghe a quelle che sono sollevate dall'adozione di qualsiasi tipo di tecnologia nel campo della cura e promozione della salute umana: la questione del bilanciamento dello sviluppo scientifico e dell'innovazione tecnologica con il rispetto della dignità della vita umana, dal concepimento alla morte naturale e dei suoi diritti fondamentali.

La problematica specifica dei sistemi di IA nel settore sanitario è indotta da una distorsione dovuta alla prospettiva del riduzionismo scientifico, che annulla la dimensione personale nella relazione di cura e prospetta di sostituire il medico e le sue attività, l'esame obiettivo e le sue valutazioni in scienza e coscienza[113], con algoritmi seppur sofisticatissimi.

La relazione di cura è un problema di fiducia[114] oltre che di comprensione intellettuale della patologia, e non si può avere fiducia in una sequenza algoritmica, della quale neanche si conosce il procedimento argomentativo[115].

In una prospettiva biogiuridica, è indispensabile che tutti i sistemi di IA, che si collocano nelle diverse fasi del procedimento clinico o chirurgico, lascino spazio ad una ponderazione dell'uomo-medico in tutte le fasi decisionali del processo, anche al fine di garantire il consenso informato[116] al paziente sulla base di un'autentica

110. Si veda il comma 2 dell'articolo 9.

111. Sul punto cfr. Messina 2023, pp. 248-253.

112. Goldberg–Adams–Blumenthal et al. 2024.

113. Cfr. Melina 2018.

114. Powell 2019.

115. Feldman–Aldana–Stein 2019.

116. Daverio 2023.

comprensione della prescrizione o terapia che gli viene proposta[117].

L'esercizio della professione medico-sanitaria è incompatibile con l'automatizzazione assoluta delle decisioni e la delega in bianco delle determinazioni, sui trattamenti da adottare o sugli interventi da eseguire, ad un calcolatore elettronico. Il medico e tutti coloro, che sono coinvolti nelle relazioni di cura e di assistenza, non possono abdicare alla responsabilità[118] delle scelte, che sono allo stesso tempo scientifiche e prudenziali, nei confronti di un sistema di IA prodotto dall'uomo.

Come sottolineato già dai bioeticisti, al fine di tutelare la dignità personale dell'uomo, sia del paziente che del medico, è necessario rimettere al centro della prassi e della normazione la persona umana[119] e la sua libertà, senza cedere il controllo ad un altro timoniere[120].

Riferimenti bibliografici

S.M. Alhashmi, I.A.T. Hashem, I. Al-Qudah (2024), *Artificial Intelligence applications in healthcare: A bibliometric and topic model-based analysis*, in "Intelligent Systems with Application", vol. 21, 2024, n. 3

A.C. Amato Mangiameli (2023), *Intelligenza artificiale*, in A.C. Amato Mangiameli, G. Saraceni (a cura di), "Cento e una voce di informatica giuridica", Giappichelli, 2023

A.C. Amato Mangiameli (2022), *Intelligenza artificiale*, in E. Sgreccia, A. Tarantino (diretto da), "Enciclopedia di Bioetica e Scienza Giuridica", Aggiornamento vol. I, Edizioni Scientifiche Italiane, 2022

A.C. Amato Mangiameli (2022-a), *Intelligenza artificiale, big data e nuovi diritti*, in "Rivista italiana di informatica e diritto", 2022, n. 1

H. Arendt (2004), *Responsabilità e giudizio*, trad. it., Einaudi, 2004

J. Bali, R. Garg, R.T. Bali (2019), *Artificial Intelligence (AI) in healthcare and biomedical research: Why a strong computational/AI bioethics framework is required?*, in "Indian Journal of Ophthalmology", vol. 67, 2019, n. 1

A. Barracca, S. Ledda, G. Mancosu et al. (2017), *La medicina fra intelligenza artificiale e necessità di un nuovo Umanesimo*, in "Giornale di clinica nefrologica e dialisi", vol. 29, 2017, n. 3

T. Bayne, A. Cleeremans, P. Wilken (eds.) (2009), *The Oxford Companion to Consciousness*, Oxford University Press, 2009

R.M. Bellino (2015), *Medicina*, in E. Sgreccia, A. Tarantino (diretto da), "Dizionario di Bioetica e Scienza Giuridica", vol. VIII, Edizioni Scientifiche Italiane, 2015

A. Berlejung, C. Frevel (a cura di) (2009), *I concetti teologici fondamentali dell'Antico e del Nuovo Testamento*, trad. it., Queriniana, 2009

117. Messinetti 2019.

118. Come ricordava Hannah Arendt, abdicare alla propria coscienza ed alla facoltà di giudizio in favore di un potere esterno o del parere già formulato dalla maggioranza è una tentazione ricorrente, soprattutto sotto i regimi totalitari: «Tante sono le ragioni per le quali il problema del diritto e della capacità di giudizio investe questioni morali di primissima importanza … Primo, come posso distinguere ciò che è giusto da ciò che è sbagliato quando la maggioranza o la totalità delle persone che mi stanno accanto ha già formulato un giudizio? … E in effetti esiste, nella nostra società, una diffusa paura di giudicare, una paura che non ha nulla a che fare con il detto biblico "Non giudicare se non vuoi essere giudicato" … In realtà, dietro il non volere giudicare si cela il dubbio che nessuno sia libero, il dubbio che nessuno sia responsabile o possa rispondere degli atti che ha commesso», cfr. Arendt 2004, pp. 16-17.

119. Palazzani 2022.

120. Κυβερνήτης = timoniere, in greco.

G. Briganti, O. Le Moine (2020), *Artificial Intelligence in Medicine: Today and Tomorrow*, in "Frontiers in Medicine", vol. 7, 2020

C. Campiglio (2023), *Giurisdizione e legge applicabile in materia di responsabilità medica (ovvero, a proposito di conflitti di qualificazioni)*, in "Rivista di diritto internazionale privato e processuale", 2023, n. 4

G. Cassano, S. Previti (a cura di) (2020), *Il diritto di Internet nell'era digitale*, Giuffrè, 2020

G. Cerrina Feroni (2023), *Intelligenza artificiale e sistemi di scoring sociale. Tra distopia e realtà*, in "Diritto dell'Informazione e dell'informatica", 2023, n. 1

M. Chary, E.W. Boyer, M.M. Burns (2021), *Diagnosis of acute poisoning using explainable artificial intelligence*, in "Computers in Biology and Medicine", vol. 134, 2021

E. Colletti (2021), *Intelligenza artificiale e attività sanitaria. Profili giuridici dell'utilizzo della robotica in medicina*, in "Rivista di diritto dell'economia, dei trasporti e dell'ambiente", vol. XIX, 2021

Comitato Nazionale per la Bioetica (2020), *Intelligenza artificiale e medicina: aspetti etici*, 29 maggio 2020

Consiglio Superiore di Sanità - Ministero della Salute (2023), *Linee guida, percorso diagnostico-terapeutico-assistenziale (PDTA), sviluppi e prospettive della Chirurgia robotica nel SSN e riflessioni sul Knowledge Transfer dell'Intelligenza Artificiale (IA)*, 13 febbraio 2023

Consiglio Superiore di Sanità - Ministero della Salute (2021), *I sistemi di intelligenza artificiale come strumento di supporto alla diagnostica*, 9 novembre 2021

J. Cowburn (2013), *Scientism: a word we need*, Mosaic Press, 2013

A. D'Adda (2022), *Danni «da robot» (specie in ambito sanitario) e pluralità di responsabili tra sistema della responsabilità civile ed iniziative di diritto europeo*, in "Rivista di diritto civile", 2022, n. 3

C. D'Elia (2023), *Gli strumenti di intelligenza artificiale generativa nel contesto sanitario: problemi di ottimizzazione delle risorse e questioni di spiegabilità*, in "Rivista italiana di medicina legale e del diritto in campo sanitario", 2023, n. 1

M. Daverio (2023), *Informed Consent and Medical Decision-Making Applying Artificial Intelligence. Ethical Aspects and Implications for the Doctor-Patient Relationship*, in "Rivista di filosofia del diritto", 2023, n. 2

M. Daverio, F. Macioce (2023), *Intelligenza artificiale e diritto alla salute nella regolazione europea: aspetti emergenti riguardo alla relazione medico-paziente*, in "Teoria e Critica della Regolazione Sociale", 2023, n. 1

A.J. Deutsch, E. Ahlqvist, M.S. Udler (2022), *Phenotypic and genetic classification of diabetes*, in "Diabetologia", vol. 65, n. 2022

European Group on Ethics in Science and New Technologies (2018), *Statement on artificial intelligence, robotics and 'autonomous' systems*, Brussels, 9 March 2018

R.C. Feldman, E. Aldana, K. Stein (2019), *Artificial Intelligence in the Health Care Space: How We Can Trust What We Cannot Know*, in "Stanford Law & Policy Review", vol. 30, 2019

E.A. Ferioli (2019), *L'intelligenza artificiale nei servizi sociali e sanitari: una nuova sfida al ruolo delle istituzioni pubbliche nel welfare italiano?*, in "BioLaw Journal – Rivista di BioDiritto", 2019, n. 3

G. Finocchiaro (2022), *La regolazione dell'intelligenza artificiale*, in "Rivista trimestrale di diritto pubblico", 2022, n. 3

V. Franceschelli (2023), *Sull'intelligenza artificiale (IA o AI)*, in "Rivista di diritto industriale", 2023, n. 1

Francesco (2022), *Discorso del Santo Padre Francesco ai membri della Federazione delle Associazioni Familiari Cattoliche in Europa*, 10 giugno 2022

C. Frevel (2009), *Cuore*, in A. Berlejung, C. Frevel (a cura di), "I concetti teologici fondamentali dell'Antico e del Nuovo Testamento", trad. it., Queriniana, 2009

E. Gabrielli, U. Ruffolo (2019), *Introduzione* alla sezione "Intelligenza Artificiale e diritto", in "Giurisprudenza italiana", 2019, n. 7

S. Gallagher, D. Zahavi (2009), *La mente fenomenologica*, Raffaello Cortina Editore, 2009

M. Gandolfini (2013), *I volti della coscienza*, Cantagalli, 2013

C.B. Goldberg, L. Adams, D. Blumenthal et al. (2024), *To do no harm – and the most good – with AI in health care*, in "Nature Medicine", vol. 30, 2024

A.G. Grasso (2023), *Diagnosi algoritmica errata e responsabilità medica*, in "Rivista di diritto civile", 2023, n. 2

J. Hatherley, R. Sparrow, M. Howard (2023), *The Virtue of Interpretable Medical AI*, in "Cambridge Quarterly of Healthcare Ethics", 2023

J. He, S.L. Baxter, J. Xu et al. (2019), *The practical implementation of artificial intelligence technologies in medicine*, in "Nature medicine", vol. 25, 2019, n. 1

P.B. Helzel (2021), *La vulnerabilità del rapporto medico-paziente nell'era delle tecnologie emergenti*, in "L'Ircocervo", vol. 20, 2021, n. 2

High-Level Expert Group on Artificial Intelligence (2020), *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*, Publications Office of the European Union, 2020

High-Level Expert Group on Artificial Intelligence (2019), *Ethics guidelines for trustworthy AI*, Publications Office of the European Union, 2019.

High-Level Expert Group on Artificial Intelligence (2019-a), *Policy and investment recommendations for trustworthy AI*, Publications Office of the European Union, 2019

J. Iqbal, D.C. Cortés Jaimes, P. Makineni et al. (2023), *Reimagining Healthcare: Unleashing the Power of Artificial Intelligence in Medicine*, in "Curēus", vol. 15, 2023, n. 9

Joint Research Centre (2020), *AI Watch. Artificial Intelligence in Medicine and Healthcare: applications availability and societal impact*, Publications Office of the European Union, 2020

S.M. Kampowski (2010), *Una libertà più grande: la biotecnologia, l'amore e il destino umano*, Cantagalli, 2010

R. Khera, A.J. Butte, M. Berkwits et al. (2023), *AI in Medicine – JAMA's Focus on Clinical Outcomes, Patient-Centered Care, Quality and Equity*, in "Journal of the American Medical Association", vol. 330 2023, n. 9

E. Lepore, A. Pylyshyn (eds.) (1999), *What is cognitive science?*, Blackwell, 1999

L.M. Lucarelli Tonini (2023), *L'IA tra trasparenza e nuovi profili di responsabilità: la nuova proposta di "AI liability directive"*, in "Il Diritto dell'informazione e dell'informatica", 2023, n. 2

B. Maier, W.A. Shibles (2011), *The Philosophy and Practice of Medicine and Bioethics. A Naturalistic-Humanistic Approach*, Springer, 2011

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

L. Melina (2018), *Coscienza e prudenza*, Cantagalli, 2018

D. Messina (2023), *La tutela della dignità nell'era digitale*, Edizioni Scientifiche Italiane, 2023

R. Messinetti (2019), *La tutela della persona umana versus l'intelligenza artificiale. Potere decisionale dell'apparato tecnologico e diritto alla spiegazione della decisione automatizzata*, in "Contratto e impresa", 2019, n. 3

A. Mishra, J. Savulescu, A. Giubilini (2021), *The Ethics of Medical AI*, in C. Véliz (ed.), "The Oxford Handbook of Digital Ethics", Oxford University Press, 2021

B. Mittelstadt (2022), *The impact of Artificial Intelligence in the doctor-patient relationship*, European Council, 2022

D. Morana, T. Balduzzi, F. Morganti (2022), *La salute "intelligente": eHealth, consenso informato e principio di non-discriminazione*, in "federalismi.it", 2022, n. 34

N. Musacchio, G. Guaita, A. Ozzello et al. (2018), *Intelligenza Artificiale e Big Data in ambito medico: prospettive, opportunità, criticità*, in "JAMD – The Journal of AMD", vol. 21, 2018, n. 3

A. Noë (2010), *Perché non siamo il nostro cervello: una teoria radicale della coscienza*, Raffaello Cortina Editore, 2010

L. Palazzani (2022), *Technological humanism: a philosophical framework for education*, in "Archivio giuridico", 2022, n. 3

L. Palazzani (2021), *AI and Health: Ethical Aspects for Regulation*, in "Teoria e critica della regolazione sociale", 2021, n. 1

L. Palazzani, S. Zannotti (a cura di) (2013), *Il diritto delle neuroscienze. Non "siamo" i nostri cervelli*, Giappichelli, 2013

A.-D.E. Parekh, O.A. Shaikh, Simran et al. (2023), *Artificial intelligence (AI) in personalized medicine: AI-generated personalized therapy regimens based on genetic and medical history: short communication*, in "Annals of Medicine and Surgery", vol. 85, 2023, n. 11

A. Pascual Leone, A. Amedi, F. Fregni, L.B. Merabet (2005), *The plastic human brain cortex*, in "Annual Review of Neuroscience", vol. 28, 2005

K. Piliuk, S. Tomforde (2023), *Artificial Intelligence in Emergency Medicine. A Systematic Literature Review*, in "International Journal of Medical Informatics", vol. 180, 2023, n. 6

E.O. Popa, M. van Hilten, E. Oosterkamp, M.-J. Bogaardt (2021), *The use of digital twins in healthcare: socio-ethical benefits and socio-ethical risks*, in "Life Sciences, Society and Policy", vol. 17, 2021, n. 6

J. Powell (2019), *Trust Me, I'm a Chatbot: How Artificial Intelligence in Health Care Fails the Turing Test*, in "Journal of Medical Internet Research", vol. 21, 2019, n. 10

Z. Pylyshyn (1999), *What's in Your Mind?*, in E. Lepore, Z. Pylyshyn (eds.), "What is cognitive science?", Blackwell, 1999

U. Ruffolo (2021), *L'Intelligenza artificiale in sanità: dispositivi medici, responsabilità e "potenziamento"*, in "Giurisprudenza italiana", 2021, n. 2

U. Ruffolo (2021-a), *Le "mobili frontiere" della responsabilità medica*, in "Giurisprudenza italiana", 2021, n. 2

S. Russell, P. Norvig (2021), *Intelligenza artificiale. Un approccio moderno*, Pearson, 2021

G. Salito (2022), *La responsabilità da algoritmo tra (teoria della) finzione e realtà sanitaria: una nuova declinazione della responsabilità medica?*, in "Rivista italiana di medicina legale e del diritto in campo sanitario", 2022, n. 4

V. Salvatore (a cura di) (2023), *Digitalizzazione, intelligenza artificiale e tutela della salute nell'Unione europea*, Giappichelli, 2023

G. Sartor (2022), *L'intelligenza artificiale e il diritto*, Giappichelli, 2022

G. Sartor (2002), *Gli agenti software: nuovi soggetti del ciberdiritto?*, in "Contratto e impresa", vol. 18, 2002, n. 2

J.R. Searle (1984), *Menti, cervelli e programmi*, trad. it., Clup-Clued, 1984

Y. Song, B. Millidge, T. Salvatori et al. (2024), *Inferring neural activity before plasticity as a foundation for learning beyond backpropagation*, in "Nature Neuroscience", vol. 27, 2024

R. Spaemann (2013), *Essere persone*, La Scuola, 2013

G. Taddei Elmi (2020), *Soggettività e responsabilità dei sistemi di IA*, in G. Cassano, S. Previti (a cura di), "Il diritto di Internet nell'era digitale", Giuffrè, 2020

Q. Teng, Z. Liu, Y. Song et al. (2022), *A survey on the interpretability of deep learning in medical diagnosis*, in "Multimedia System", vol. 28, 2022

A.E. Tozzi (2023), *Verso una leadership clinica dell'intelligenza artificiale per la salute*, in "L'Endocrinologo", vol. 24, 2023, n. 5

A.M. Turing (1950), *Computing machinery and intelligence*, in "Mind", vol. 59, 1950, n. 236 (disponibile nella traduzione italiana a cura di G. Lolli, A.M. Turing, *Intelligenza meccanica*, trad. it., Bollati Boringhieri, Torino 1994)

UNESCO (2021), *Recommendation on Ethics of Artificial Intelligence*, 23rd November 2021

UNESCO (2017), *Report of the International Bioethics Committee on Big Data and Health*, 15th September 2017

UNESCO (2005), *Universal Declaration on Bioethics and Human Rights*, 2005

L. Wang, R. Gao, J. Váncza et al. (2019), *Symbiotic human-robot collaborative assembly*, in "Manufacturing Technology", vol. 68, 2019

World Health Organization (2021), *Ethics and governance of artificial intelligence for health: WHO guidance*, 2021

N. Xu, D. Yang, K. Arikawa, C. Bai (2023), *Application of artificial intelligence in modern medicine*, in "Clinical eHealth", vol. 6, 2023, n. 12