LORELLA MEOLA
INTELLIGENZA ARTIFICIALE E RELAZIONE
MEDICO-PAZIENTE:
IMPLICAZIONI EPISTEMICHE ED ETICHE

1. *Intelligenza artificiale per la medicina*

L'intelligenza artificiale (IA) sta trovando ampia applicazione in medicina¹, con l'ambizione di rivoluzionare tanto i paradigmi di ricerca biomedica quanto le pratiche cliniche².

In particolare, essa avanza come una delle tecnologie più promettenti all'interno della svolta segnata dalla salute digitale (*eHealth*), ovvero informazioni e servizi sanitari forniti e progressivamente migliorati tramite internet e le tecnologie correlate³. Applicazioni per la salute, dispositivi indossabili, cartella clinica elettronica, telemedicina, ma ancora assistenti personali digitali, *chatbot* sono solo alcuni esempi della trasformazione digitale in corso⁴ e sono sintomatici di come tale mutamento stia muovendo in una triplice direzione: maggiori centralità e autonomia del paziente all'interno della relazione di cura, incremento esponenziale dei dati disponibili e personalizzazione del percorso terapeutico⁵.

-
- 1 J. He, S.L. Baxter, J. Xu, X. Zhou, K. Zhang, *The practical implementation of artificial intelligence technologies in medicine*, in "Nature Medicine", 25(1), 2019, pp. 30-36; B. Kallis, M. Collier, R. Fu, *10 Promising AI Applications in Health Care*, "Harvard Business Review", 2018, pp. 1- 5; F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, Y. Wang, *Artificial intelligence in health-care: past, present and future*, in "Stroke Vasc Neurol", 2(4), 2017, pp. 230-243.
 - 2 A. Blasimme, E. Vayena, *The Ethics of AI in Biomedical Research, Patient Care, and Public Health*, in M.S. Dubber, F. Pasquale, S. Das, *The Oxford Handbook of Ethics of AI*, Oxford University Press, 2020, pp. 703- 718; E. Topol, *High-performance medicine: the convergence of human and artificial intelligence*, in "Nature Medicine", 25, 2019, pp. 44-56
 - 3 G. Eysenbach, *What is e-health?*, in "J Med Internet Res", 3:e20, 2001.
 - 4 World Health organization (WHO), *Global Observatory. New Horizons for Health through mobile technologies*, Geneve: World Health Organization (WHO), 2011.
 - 5 B. Prainsack, *Personalized Medicine: Empowered Patients in the 21st Century?*, NYU Press, New York, 2017.

Non più una medicina, per così dire, “a taglia unica”, rivolta al paziente medio, basata sull’osservazione dei sintomi e che procede per tentativi ed errori, vale a dire una medicina che formula la diagnosi e prescrive il trattamento più probabili; ma una pratica precisa, rivolta alla specificità e unicità di ciascun paziente⁶. L’obiettivo è dare forma a una medicina preventiva, personalizzata e di precisione, che procede in maniera mirata, alla luce delle caratteristiche peculiari del paziente, ricostruite in maniera precisa attraverso l’acquisizione, l’analisi e l’elaborazione dei dati molecolari mediante algoritmi. Tale approccio consente di agire direttamente sul fattore patologico esatto o individuare preventivamente il trattamento adeguato per il paziente, facilitando il processo decisionale all’interno della relazione terapeutica. Si tratta di un modello di ricerca e di pratica fondato sulla centralità dei dati⁷, che tiene conto della variabilità individuale dei geni, ma anche dell’ambiente e dello stile di vita di ciascuna persona⁸. Sequenziamento del genoma umano a basso costo, ampia applicazione delle scienze -omiche⁹, test genetici, dispositivi digitali di tracciamento, diagnostica per immagini sono alcuni esempi di come le biotecnologie avanzate stiano producendo una massiccia quantità di dati eterogenei¹⁰, che difficilmente il singolo ricercatore o medico possono raccogliere, contenere e valutare. Il riferimento ai *Big Data* richiede una notevole capacità di calcolo, che è propria delle tecnologie più avanzate e tra le quali l’IA è senza dubbio quella più promettente: data la capacità di contenimento, scansione e dunque controllo dei dati, i sistemi di IA, con il tempo, potrebbero diventare più informati e accurati dei medici e questi ultimi potrebbero non essere più in grado di rinunciare agli algoritmi¹¹.

6 X. Guchet, *La médecine personnalisée. Un essai philosophique*, Le Belles Lettres, Paris, 2016.

7 S. Leonelli, *Data-Centric Biology. A Philosophical Study*, University Chicago Press, 2016.

8 *The Precision Medicine Initiative*, in <<https://obamawhitehouse.archives.gov/precision-medicine>>, (ultimo accesso 30 giugno 2022).

9 Le scienze -omiche sono discipline biomolecolari, volte a stabilire le correlazioni tra le caratteristiche proprie del fenotipo e la quantificazione delle molecole biologiche coinvolte nella struttura, funzione, dinamica di una cellula, un tessuto, un organismo; cfr., European Commission, *Commission staff working document: use of “-omics” technologies in the development of personalized medicine*, Brussels, 2013.

10 Nuffield Council on Bioethics, *The Collection, linking and use of data in biomedical research and health care: ethical issues*; 2015.

11 E.J. Topol, *High-performance medicine: the convergence of human and artificial intelligence*, cit.

Sebbene siamo ancora all'alba delle applicazioni di IA in medicina, tuttavia c'è motivo di pensare che, in futuro, gli algoritmi supereranno le prestazioni dei medici in termini di velocità, precisione e affidabilità¹². Inoltre, se la personalizzazione è l'obiettivo da perseguire¹³, si può sostenere che non può esistere medicina personalizzata senza IA¹⁴. Insomma, la medicina del futuro è inestricabilmente connessa alle sorti dell'IA.

Quest'ultima potrebbe avere un impatto notevole sulla medicina, in quanto può integrare o subentrare alle pratiche mediche tradizionali: le macchine potrebbero non semplicemente supportare l'attività del medico, ma addirittura sostituire la figura del medico e riscrivere il ruolo del paziente¹⁵; pertanto diviene urgente affrontare le questioni etiche da essa sollevate e, in particolare, interrogarsi sulle implicazioni che tali sistemi avranno sulla relazione di cura¹⁶. Se l'IA mostra capacità diagnostiche e terapeutiche superiori rispetto a quelle dei medici, si potrebbe pensare che questi ultimi abbiano un obbligo epistemico nei confronti delle macchine nel corso del processo decisionale. D'altro lato, però, occorre tener conto del fatto che tale obbligo potrebbe confliggere con altri valori propri della relazione terapeutica, come per esempio l'empatia che si instaura tra medico e paziente, e che, pertanto, potrebbero fornire un motivo valido per rinunciare all'obbligo epistemico del medico verso la macchina¹⁷. Tale potenziale conflitto deve sollecitare una riflessione critica sul ruolo che l'IA deve avere nel processo decisionale all'interno della relazione medico-paziente.

2. *Il problema della scatola nera*

Come ricordato sopra, i fattori che maggiormente hanno contribuito agli sviluppi dell'IA sono la disponibilità di una massiccia quantità di dati, non-

12 Y. Jiang et alii, *Artificial intelligence in healthcare: past, present and future*, cit.

13 N. Rose, *Personalized Medicine: Promises, Problems and Perils of a new Paradigm for Healthcare*, in «Procedia. Social and Behavioral Sciences», vol. 77, 2013, pp. 341- 352.

14 B. Mesko, *The Role of Artificial Intelligence in Precision Medicine*, in “*Expert Review of Precision Medicine and Drug Development*”, 5, 2017, pp.239- 241.

15 H. ten Have, B. Gordijn, *Medicine and Machines*, in “*Medicine, Health Care and Philosophy*”, 25, 2022, pp. 165- 166.

16 CNB, CNBBSV, *Intelligenza artificiale e medicina: aspetti etici*, 29 maggio 2020, in <https://bioetica.governo.it/media/4260/p6_r_2020_gm_intelligenza-artificiale-it.pdf>, (ultimo accesso 20 giugno 2022).

17 F. Cabitza, R. Rasoini, G.F. Gensini, *Unintended Consequences of Machine Learning in Medicine*, in “*JAMA*”, 318 (6): 517–518, 2017.

ché la sofisticata capacità di elaborazione degli algoritmi. In particolare, il sistema di *machine learning*, sottogruppo dell'IA, è un approccio di apprendimento automatico, che consente alla macchina di individuare, mediante un confronto veloce e dinamico, associazioni tra i dati. Attraverso modelli matematici, esso permette a un sistema di computer di imparare senza ricevere istruzioni dirette e continuare a migliorare nell'apprendimento in virtù della sua esperienza e dunque senza la guida del programmatore. Giacché il fine dell'IA è simulare le attività cognitive umane, quali, per esempio, apprendimento, risoluzione dei problemi, assunzione di decisioni, è possibile impiegare una rete neurale, vale a dire una serie di algoritmi modellati in base al cervello umano: in quest'ultimo caso si parla di *deep learning*¹⁸.

L'ingente mole di dati che alimenta tali sistemi offusca sia la trasparenza sia la prevedibilità del loro funzionamento e conferisce a molti sistemi di IA un alto livello di opacità¹⁹. Quest'ultima è intesa come imperscrutabilità del meccanismo soggiacente ai processi decisionali degli algoritmi e dunque mancanza di chiarezza dei passaggi attraverso i quali si interpretano i dati²⁰. A partire da dati impiegati come *input*, si ottengono degli *output*, senza tuttavia disporre di una spiegazione del come e perché sono stati prodotti determinati contenuti, raccomandazioni, decisioni. L'opacità può essere relativa e pertanto interessare solo l'utente finale, oppure può riguardare sia l'utente sia il programmatore. Ciò rende il processo attraverso il quale l'IA produce i suoi *output* paragonabile a una scatola nera – *black box*²¹.

Pertanto, una delle sfide epistemiche ed etiche più audaci poste dall'IA riguarda l'"esplicabilità". Vale a dire rendere conto dei meccanismi di elaborazione e interpretazione dei dati e poter così consentire agli umani di valutare l'appropriatezza o l'inadeguatezza dei risultati prodotti.

18 B. Mondal, *Artificial Intelligence: State of the Art*, in V.E. Balas et alii, *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, Springer, 2020, pp.389-425.

19 B. Heinrichs, S.B. Eickhoff, *Your Evidence? Machine Learning algorithms for medical diagnosis and prediction*, in "Human Brain Mapping", 41, 2020, pp. 1435- 1444.

20 J. Burrel, *How the machine 'thinks': Understanding Opacity in machine learning algorithms*, in "Big Data Society", 2016, <<https://doi.org/10.1177/2053951715622512>>; T. Zarsky, *The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making*, in "Science, Technology and Human Values", 41 (1), 2016, pp. 118–132.

21 M. Carabantes, *Black-box artificial intelligence: an epistemological and critical analysis*, in "AI & Society", 35, 2020, pp. 309-317.

Nel campo specifico della medicina, tale difficoltà si inasprisce, se si tiene conto che i sistemi di IA possono generare dei pregiudizi, i quali, a loro volta, potrebbero condurre a discriminazioni²²: l'efficacia di un trattamento potrebbe essere stabilita secondo canoni parziali, senza tener conto in maniera adeguata della specificità del paziente, le cui esigenze peculiari -personali, diagnostiche e terapeutiche- potrebbero risultare sfavorite. Inoltre, i sistemi di IA possono incorporare dei pregiudizi propri del processo decisionale umano -come quando i medici valutano i casi di specie tenendo conto dell'età del paziente- e applicarli puntualmente, anche quando tali criteri potrebbero alterare la correttezza della valutazione clinica specifica. In terzo luogo, i sistemi di IA formulano le loro decisioni sulla base di gruppi di pazienti e potrebbero non essere sempre in grado di scegliere il miglior trattamento per un determinato paziente.

L'“esplicabilità” dei meccanismi decisionali degli algoritmi potrebbe correggere gli errori di un modello, ottimizzare le *performances* e costruire un quadro etico e giuridico rassicurante²³. A tal riguardo, la Commissione Europea ha proposto il *Libro Bianco sull'intelligenza artificiale –Un approccio europeo all'eccellenza e alla fiducia*, individuando proprio nell'esplicabilità uno dei requisiti fondamentali, finalizzato alla promozione della fiducia degli utenti nei confronti dei sistemi di IA²⁴.

Allo stesso modo, nel documento *Orientamenti etici per una IA affidabile* si afferma il diritto degli utenti a una spiegazione adeguata del processo di decisione ogni volta che l'IA influisce considerevolmente sulla vita delle persone²⁵.

Pur non avendo raggiunto una definizione condivisa, tuttavia numerosi sono i tentativi messi in campo al fine di individuare il perimetro semantico e applicativo dell'esplicabilità²⁶. I *Principi di Asilomar* parlano di traspa-

22 B.D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, L. Floridi, *The Ethics of Algorithms: Mapping the Debate*, in “*Big Data & Society*”, 3, 2, 2016, pp. 1-21.

23 L. Floridi et alii, *AI4People –an ethical framework for a good AI society: opportunities, risks, principles, and recommendations*, cit.

24 Commissione Europea, *Libro bianco sull'intelligenza artificiale – Un approccio europeo all'eccellenza e alla fiducia*, in < https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_it.pdf>, (ultimo accesso 20 giugno 2022).

25 Commissione europea, Direzione generale delle Reti di comunicazione, dei contenuti e delle tecnologie, *Orientamenti etici per un'IA affidabile*, Ufficio delle pubblicazioni, 2019, <<https://data.europa.eu/doi/10.2759/640340>>, (ultimo accesso 20 giugno 2022)

26 A. Adadi, M. Berrada, *Peeking Inside the Black-Box: a Survey on Explainable Artificial Intelligence (XAI)*, IEEE Access, 6, 2018, pp. 52138-52160.

renza²⁷; in termini di responsabilità si pronuncia, dal canto suo, il *Gruppo Europeo di etica delle scienze e delle nuove tecnologie*²⁸; si fa invece riferimento sia alla trasparenza sia alla responsabilità nei principi generali della seconda versione di *Ethically Aligned Design, dell'Institute of Electrical and Electronics Engineers (IEEE)*²⁹; inoltre, il *Comitato per l'intelligenza artificiale della Camera dei Lord del Regno Unito* parla di decifrabilità³⁰.

Luciano Floridi e colleghi sintetizzano tale pluralità linguistica, riducendola al suo significato epistemologico di intelligibilità, in quanto il richiamo all'esplicabilità deve rispondere in primo luogo alla domanda: «come funziona un processo algoritmico decisionale?», nonché al suo senso etico di responsabilità, in quanto l'esplicabilità deve rispondere anche alla domanda fondamentale: «chi è responsabile del modo in cui tale sistema funziona?». Dal punto di vista epistemico una chiara definizione di esplicabilità consente un attento esame delle decisioni assunte da un sistema: promette a coloro che programmano, sviluppano e regolano i sistemi di IA, nonché agli utenti finali di individuare errori sistematici, correggere decisioni errate e migliorare le prestazioni del sistema. Dal punto di vista etico, risulta garantita l'autonomia degli individui, vengono correttamente imputate le responsabilità, sono evitati pregiudizi discriminatori ed è così promossa la fiducia degli utenti verso i sistemi di IA³¹. Nel sostenere l'imprescindibilità di tale valore, Floridi declina quest'ultimo in termini di principio e lo colloca nel quadro del *principlismo biomedico* proposto da James Childress e Tom Beauchamp³², in quanto, tra tutti i campi di applicazione dell'etica, la bioetica sembra emergere come la disciplina che

27 *Asilomar AI Principles*, adottati a conclusione della Asilomar Conference on Beneficial AI, California, 2017.

28 European Group on Ethics in Science and New Technologies, *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*, 2018, in <https://ec.europa.eu/info/news/ethics-artificial-intelligence-statement-ege-released-2018-apr-24_en>, (ultimo accesso 24 giugno 2022).

29 The IEEE Initiative on Ethics of Autonomous and Intelligent Systems, *Ethically Aligned Design*, v2. IEEE, 2017, in <<https://ethicsinaction.ieee.org/wp-content/uploads/ead1e.pdf>>, (ultimo accesso 23 giugno 2022).

30 House of Lords Artificial Intelligence Committee, *AI in the UK: ready, willing and able?* Retrieved September 18, 2018, <<https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm>>, (ultimo accesso 23 giugno 2022).

31 W.K. Diprose, N. Buist, N. Hua, Q. Thurier, G. Shand, R. Robinson, *Physician understanding explainability, and trust in a hypothetical machine learning risk calculator*, in "*Journal of the American Medical Informatics Association*", 2784, 2020, pp. 592-600.

32 T.L. Beauchamp, J.F. Childress, *Principi di etica biomedica*, Le Lettere, Firenze, 1999.

più si avvicina all'etica digitale e meglio chiarisce ruoli e responsabilità degli agenti all'interno del nuovo ecosistema tecnologico³³. Seguendo questo schema, egli sostiene innanzitutto che l'IA deve essere benefica e non malefica, vale a dire una riflessione sui sistemi di IA richiede innanzitutto di valutare il bene o il male che essi stanno effettivamente producendo all'interno delle società e le modalità attraverso le quali lo fanno; in secondo luogo, l'IA deve promuovere e non limitare l'autonomia umana, confrontando ogni volta le possibili modalità di azione umane con quelle dell'IA; e ancora l'IA deve essere giusta, vale a dire che la tecnologia o, più precisamente, le persone e le organizzazioni che la sviluppano e la implementano sono ritenute responsabili di un eventuale risultato negativo e dei motivi che lo hanno determinato. Più in generale, sostiene Floridi, dobbiamo inquadrare il rapporto tra umani e tecnologia su basi facilmente comprensibili³⁴. Pertanto, ai quattro capisaldi mutuati dall'etica biomedica, egli aggiunge il principio di esplicitabilità, dunque intelligibilità dei processi algoritmici e netta individuazione delle responsabilità rispetto alle decisioni prodotte, facendo di tale principio il riferimento precipuo di un quadro etico e giuridico all'interno del quale governare gli sviluppi delle nuove tecnologie.

Fare chiarezza all'interno della scatola nera permette di equilibrare il rapporto tra umani e macchine, vale a dire tra il potere decisionale che riserviamo a noi stessi e quello che deleghiamo ad agenti artificiali³⁵. Al riguardo, la *Dichiarazione di Montreal* ribadisce la necessità di un rapporto proporzionato tra il processo deliberativo guidato dall'uomo e quello svolto dalla macchina, ponendo come obiettivo dello sviluppo dell'IA la promozione dell'autonomia di tutti gli esseri umani e il controllo sui processi decisionali delle tecnologie³⁶. Il *Gruppo Europeo dell'Etica delle scienze e delle nuove tecnologie* sostiene che i sistemi autonomi non devono compromettere la libertà degli umani, riservando a questi ultimi la facoltà di definire in autonomia le norme e i canoni secondo i quali vivere³⁷. I *Principi*

33 L. Floridi (eds.), *Ethics, Governance, and Policies in Artificial Intelligence*, Springer, 2021.

34 L. Floridi et alii, *ai4People – an ethical framework for a good AI society: opportunities, risks, principles, and recommendations*, cit.

35 G. Tamburrini, *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Carocci, Roma, 2020.

36 *Dichiarazione di Montréal sullo sviluppo responsabile dell'intelligenza artificiale*, 2018.

37 Commissione europea, Direzione generale della Ricerca e dell'innovazione, Gruppo europeo sull'etica nelle scienze e nelle nuove tecnologie, *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*, Bruxelles 2018.

di *Asilomar*, allo stesso modo, dichiarano l'imprescindibilità del principio di autonomia, sostenendo che tocca agli esseri umani scegliere le modalità e le forme di delega delle decisioni ai sistemi di IA, al fine di raggiungere obiettivi posti già e sempre dall'uomo.

In sintesi, tutti questi documenti tracciano i confini di autonomia e responsabilità di umani e macchine, bandendo il rischio di cedere un margine decisionale troppo ampio alle macchine e promuovendo l'autonomia decisionale degli uomini, al fine di ridurre i rischi e tutelare la sicurezza per gli uomini, nonché implementare l'efficacia di tali sistemi.

3. Opacità e relazione medico-paziente

L'opacità che caratterizza i sistemi di IA potrebbe causare evidentemente delle conseguenze rispetto al processo decisionale che prende forma all'interno della relazione terapeutica. Nella misura in cui i meccanismi che producono specifici risultati medici non risultano intelligibili e dunque il medico non è in grado di fornire al paziente le informazioni necessarie per sostenere quest'ultimo nel percorso deliberativo, si strutturerebbe una «medicina della scatola nera», che potrebbe minare le basi della responsabilità del medico e dell'autonomia del paziente³⁸. Anche quando ci imbattiamo in risultati eccellenti raggiunti dalle macchine, senza aver però compreso le modalità e le ragioni di certi risultati, rischiamo di perdere quello che è ormai considerato un cardine della relazione terapeutica, vale a dire il consenso informato e la relativa consapevolezza dell'analisi rischio-beneficio di un determinato trattamento. Pertanto, la problematica intelligibilità dovrebbe costituire una ragione valida per concedere ai pazienti il diritto di rifiutare un trattamento proposto dall'IA e il diritto di richiedere un trattamento alternativo³⁹.

A partire dagli anni Sessanta e Settanta del secolo scorso, il rispetto dell'autonomia del paziente è assunto a principio morale imprescindibile della pratica medica⁴⁰. La rivoluzione digitale dell'ultimo decennio, a sua

38 W.N. Price, *Medical Malpractice and Black-Box Medicine*, in I. Cohen, H. Lynch, E. Vayena, U. Gasser (eds.), *Big Data, Health Law, and Bioethics*, Cambridge, Cambridge University Press, Cambridge, 2018, pp. 295-306.

39 T. Ploug, S. Holm, *The right to refuse diagnostics and treatment planning by artificial intelligence*, in "Medicine, Health Care and Philosophy", 23, 2020, pp. 107-114.

40 R.M. Veatch, *Autonomy's temporary triumph*, in "Hastings Center Report", 14, 5, 1984, pp. 38-40.

volta, ha contribuito a fortificarlo, fornendo informazioni e strumenti necessari ai pazienti affinché questi ultimi acquisiscano maggiore consapevolezza e potere decisionale sulla propria malattia e salute⁴¹.

Sebbene il principio di autonomia sia suscettibile di molteplici interpretazioni⁴², tuttavia esso può essere inteso, innanzitutto, come la capacità dell'individuo di assumere decisioni in base a un progetto di vita che ciascuno sceglie per sé, ma anche come la capacità di agire in conformità a tale progetto, in quanto non vi sono impedimenti esterni all'esercizio della propria autonomia⁴³.

Il riconoscimento dell'autonomia del paziente nelle decisioni che riguardano la salute e la malattia ha fondato un nuovo modello della relazione tra medico e paziente come completamente alternativo a quello tradizionale e paternalistico derivato dalla medicina ippocratica. Il richiamo all'autonomia è dunque un assunto necessario per arginare i rischi del paternalismo medico legati ai progressi delle tecnologie biomediche, come, per esempio, l'abuso dei mezzi terapeutici, senza adeguatamente informare il paziente e ricevere da questi un legittimo consenso. Lo strumento in cui si concreta il principio di autonomia è il consenso informato, ovvero la pratica attraverso la quale si chiede al paziente l'approvazione all'intervento da attuare e che è ormai assunto come punto di convergenza della pratica medica quotidiana. La scelta del paziente deve essere preceduta e sostenuta dal dialogo con il medico; sarà quest'ultimo a mettere il paziente a conoscenza, in modo adeguato, della verità circa il proprio stato di salute e malattia, le ragioni che richiedono uno specifico percorso terapeutico, le prevedibili conseguenze della terapia, nonché eventuali procedure alternative⁴⁴. Il consenso informato non deriva tanto dalla mancanza di fiducia nei confronti del medico – i pazienti continuano a fidarsi dei medici e pertanto a essi continuano a rivolgersi-, quanto dal riconoscimento da parte del paziente della sovranità sul proprio corpo e sulla propria vita, il che richiede titolarità decisionale. Esso risulta anche dalla consapevolezza che il bene del paziente non è più una definizione oggettiva, ma è ora scelto e stabilito dal paziente stesso, a partire e attraverso il dialogo con il proprio medico. Sin dalla nota definizione di salute formulata dall'Organizzazione mondiale della

41 B. Prainsack, *Personalized Medicine: Empowered Patients in the 21st Century?*.

42 Per una sintesi, si veda, G. Barazzetti, *Libertà e medicina. Il principio di autonomia nell'etica biomedica*, Bruno Mondadori, Milano-Torino, 2011.

43 T. Beauchamp, J. Childress, *Principi di etica biomedica*, cit.

44 R.R. Faden, T.L. Beauchamp, *A history and theory of informed consent*, Oxford University Press, Oxford- New York, 1986.

sanità⁴⁵, l'attenzione si è spostata dalla malattia alla presa in considerazione della persona nella sua totalità e specificità, tenendo conto dei suoi valori e delle sue preferenze. Il paziente diviene parte attiva della relazione terapeutica e pertanto il medico dialogherà con il paziente, cercando di conciliare le informazioni strettamente mediche con valori, aspettative e timori personali. In tal modo, il terapeuta si assicura che il paziente abbia effettivamente compreso le informazioni ricevute e le abbia contestualizzate nel proprio progetto di vita, al fine di assumere una decisione autonoma e razionale.

Sebbene questa definizione della relazione terapeutica sia indubbiamente generica, ciò che deve emergere è la ricchezza del rapporto medico-paziente, che non è limitato alla sola malattia, ma si estende all'intera persona del paziente. Il medico non è mero depositario di informazioni oggettive e il modello terapeutico fondato sulla centralità del paziente è ricco di valori che vanno oltre la mera dimensione tecnica e scientifica, instaurando un rapporto fondato sull'ascolto, sulla comprensione e sul supporto alla persona del paziente nella sua totalità⁴⁶. Il medico deve fornire le informazioni relative alla situazione clinica e, allo stesso tempo, aiutare a chiarire i valori racchiusi nelle opzioni disponibili e stabilire perché certi valori relativi alla salute siano più degni di altri di essere perseguiti⁴⁷.

All'interno del triangolo terapeutico (paziente, medico, IA), il dialogo informativo e dunque l'autonomia e la razionalità decisionali potrebbero essere compromessi⁴⁸. Innanzitutto, i medici si rivelerebbero incapaci di fornire informazioni chiare e trasparenti al paziente, in quanto non dispongono degli elementi sufficienti per convalidare o scartare una certa soluzione offerta dai sistemi di IA, in ragione della loro opacità. Tale difficoltà si potrebbe inasprire in quanto gli algoritmi operano una analisi su gruppi o sottogruppi di pazienti che presentano caratteristiche simili, associate a determinati schemi, e pertanto potrebbero non tener conto in modo adeguato della variabilità e dunque peculiarità del singolo paziente. Tali classificazioni, operate sulla base di criteri e finalità differenti, nonché di

45 Secondo la Costituzione dell'Organizzazione Mondiale della Sanità, la salute non la sola assenza di malattia, ma «uno stato di totale benessere fisico, mentale e sociale», OMS, «Costituzione dell'organizzazione mondiale della sanità», in <<http://www.who.int/net>>.

46 N. Mead, P. Bower, *Patient-centriness: a conceptual framework and review of the empirical literature*, in "Social Science & Medicine", 51, 2000, pp. 1087- 1110.

47 E.J. Emanuel, L.L. Emanuel, *Four models of the Physician-Patient Relationship*, in "Journal of the American Medical Association", 267 (16), 1992, pp. 2221-2226.

48 B.D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, L. Floridi, *The ethics of algorithms: mapping the debate*, in "Big Data & Society", 2016, pp. 1-21.

considerazioni extra-sanitarie, quali origine etnica o genere di appartenenza, potrebbero ingenerare decisioni discriminatorie o arbitrarie⁴⁹. Inoltre, ai sistemi di IA, per quanto questi ultimi possano essere anche più precisi ed efficaci dei medici, mancherebbe la capacità etica di relazionarsi alla persona del paziente, prendendone in considerazione le preferenze e gli interessi ritenuti rilevanti al fine del trattamento. Quand'anche i sistemi di IA fossero in grado di valutare non solo dati medici, ma anche le altrettanto fondamentali informazioni non mediche, la disponibilità dei dati potrebbe essere limitata; i dati potrebbero essere non aggiornati o protetti dalla *privacy*; inoltre, potrebbero non riflettere le preferenze e gli interessi dei pazienti allo stato attuale e in relazione a una prospettiva di vita dischiusa da un certo quadro clinico.

Dati tali presupposti, la medicina rischierebbe di tendere verso l'impersonale, riducendo il paziente a una comprensione oggettiva di una categoria medico-biologica, costruita su base statistica. Ne deriverebbe una riformulazione dell'intero sistema di cura, che interessa la relazione terapeutica, la nosografia, le politiche sanitarie così come la ricerca biomedica, e che procederebbe verso una standardizzazione delle pratiche e dei soggetti coinvolti. In particolare, sembra tracciarsi una scissione tra una visione riduzionistica e una olistica del paziente, che riaprirebbe un vecchio dibattito della filosofia della medicina, in cui si contrappongono la medicina come arte e la medicina come tecnica. Senza entrare nel vivo della questione, possiamo ricordare, con Jaspers, che il problema sorge nella misura in cui questi due aspetti vengono disgiunti: l'oggetto della medicina, il malato, non è mai completamente oggettivabile; nella sua azione tecnica, la medicina poggia certamente su saperi scientifici, dati quantitativi, strumenti di misura, ma non ignora che il paziente è altra cosa che un ammasso di numeri; egli è piuttosto un'esistenza, che non si lascia mai quantificare. La medicina non è solo una scienza oggettiva, ma reclama un'arte speciale, ovvero il senso del vivente che non può mai essere pienamente razionalizzato⁵⁰.

4. *Autorità epistemica ed etica del medico*

Rispetto ai risultati prodotti dai sistemi di IA si è ingenerata una controversia nella letteratura sul tema. Da un lato, vi sono coloro che affermano

49 A. Blasimme, E. Vayena, *The Ethics of AI in Biomedical Research, Patient Care, and Public Health*, cit.

50 K. Jaspers, *Il medico nell'età della tecnica*, Raffaello Cortina, Milano, 1991.

che gli utenti debbano comprendere le ragioni che implicano esiti determinati, a sostegno della fiducia dei pazienti nei confronti di tali sistemi⁵¹; d'altro lato, vi sono coloro che ritengono l'esplicabilità un principio sopravvalutato e chiedono che sia garantito esclusivamente un certo livello di accuratezza e affidabilità, assumendo che la validità del risultato renda inutile la valutazione dei mezzi e metodi impiegati.

A quest'ultimo riguardo, Alex London sostiene che, in fondo, la decisione medica è da sempre avvolta da una certa opacità: l'eziologia di alcune patologie, nonché l'efficacia dei trattamenti, per esempio, possono risultare incerti anche senza chiamare in causa l'IA⁵². Conoscenza e pratica mediche sarebbero per lo più una sintesi di risultati empirici e tradizione clinica ereditati nel corso dei decenni; il bene del paziente rifletterebbe, da sempre, l'esperienza del beneficio senza che rispetto a quest'ultimo vi sia una conoscenza sufficiente delle cause che lo producono. In altri termini, la scienza medica ha sempre proceduto attraverso la messa alla prova di meccanismi di associazione delle evidenze empiriche, come avviene con i *big data*, senza conoscere le teorie che fondano e spiegano i funzionamenti patologici e quelli terapeutici. Ma soprattutto, tale approccio empirico si è sempre rivelato più efficace dei tentativi di riferirsi a una teoria al fine di individuare il percorso terapeutico più adatto. Per esempio, il ricorso alla mastectomia totale rispetto a soluzioni alternative meno invasive è stato a lungo sostenuto dalla teoria fisiopatologica, secondo la quale la rimozione di quanto più tessuto possibile ridurrebbe la probabilità di recidiva del cancro. Una serie di studi clinici successivi ha dimostrato, invece, la falsità di tale teoria. Questo è un esempio di come, in medicina, con il tempo le spiegazioni teoriche non solo si sono rivelate false, ma talvolta hanno addirittura provocato danni ai pazienti, mentre, al contrario, le evidenze cliniche hanno prodotto i risultati più efficaci. London conclude che sarebbe più opportuno dare priorità all'accuratezza diagnostica o predittiva di un sistema di IA, invece di fondare tali sistemi su un esigente principio di esplicabilità,

51 J.C. Bjerring, J. Busch, *Artificial Intelligence and Patient-Centered Decision-Making*, in "Philosophy & Technology", 34, 2021, pp. 349- 371; B. Heinrichs, S.B. Eickhoff, *Your Evidence? Machine Learning algorithms for medical diagnosis and prediction*, in "Human Brain Mapping", 41, 2020, pp. 1435- 1444; A. Holzinger, A. Carrington, H. Müller, *Measuring the Quality of Explanations: the System Causability Score (SCS)*, in "KI-Künstliche Intelligenz", 34, 2020, pp. 193- 198; C. Rudin, J. Radiin, *Why are We Using Black Box Models in AI When We Don't Need To?, A Lesson from an Explainable AI Competition*, in "Harvard Data Science Review", 1 (2), 2019.

52 A.J. London, *Artificial Intelligence and black-box medical decisions: accuracy versus explainability*, cit.

guadagnando così in termini di affidabilità clinica⁵³. Insomma, le macchine rivendicherebbero un livello così elevato di accuratezza e competenza da non rendere necessaria la richiesta di trasparenza del loro operato.

Al contrario, nel caso degli agenti umani che assumono decisioni e compiono azioni, esiste sostanzialmente la possibilità di interrogare il medico, comprendere il fondamento epistemico e normativo su cui egli ha costruito la propria raccomandazione e verificarne la plausibilità rispetto a criteri generali di competenza epistemica. Non si tratta di assumere la supervisione dell'agire medico da parte del paziente, ma di colmare l'inevitabile asimmetria che vi è tra le due parti con un atto di fiducia intelligente, ovvero di un sentimento di fiducia sostenuto su basi razionali, individuabili nella competenza rispetto allo specifico ambito di azione, onestà e attendibilità del medico⁵⁴. Fidarsi di un'altra persona vuol dire concederle epistemicamente e normativamente tale autorità, vale a dire significa riconoscere delle ragioni per credere in ciò che sta dicendo sulla base della precisione dimostrata nel tempo in un ambito specialistico, della coerenza nel portare a compimento il proprio impegno e della consapevolezza dei propri limiti e delle altrui possibilità⁵⁵.

Nei confronti dei pazienti, il medico è un'autorità epistemica e morale⁵⁶. In primo luogo, egli è competente rispetto a un ambito specifico e non solo in quanto possiede informazioni accurate, ma dispone anche dell'abilità di conciliare tali informazioni con la peculiarità del caso. Egli deve individuare eventuali percorsi terapeutici alternativi, segnalare possibili limiti e incertezze rispetto al caso concreto; il medico è un interprete della salute⁵⁷, vale a dire che deve mediare tra conoscenza medica da una parte e valori, interessi e preferenze del paziente dall'altra parte. La selezione dei parametri ritenuti rilevanti, la valutazione di ogni possibile obiettivo terapeutico, la scelta dei mezzi per raggiungere gli scopi individuati e le relative conseguenze per la vita del paziente sono tutti aspetti intrinsecamente carichi di natura normativa, non riducibili alle sole categorie oggettive. Il vocabolario della medicina, che raccoglie concetti quali salute, malattia, dolore, guarigione e così via

53 *Ivi*.

54 O'Neill, *Una questione di fiducia, Vita e Pensiero*, Milano, 2003.

55 O'Neill, *Accountability, Trust and Informed Consent in medical practice and research*, in "*Clin Med (Lond)*", 4 (3), 2004, pp.269-276.

56 F. Funer, *The Deception of Certainty: how Non-Interpretable Machine Learning Outcomes Challenge the Epistemic Authority of Physicians. A deliberative-relational approach*, in "*Medicine, Health Care and Philosophy*", 25, 2022, pp. 167- 178.

57 F. Svenaeus, *Phenomenological Bioethics. Medical Technologies, Human Suffering, and the Meaning of Being Alive*, Routledge, 108.

pone una questione etica oltre che epistemica, in cui la dimensione oggettiva va integrata con una valutazione della prospettiva soggettiva, e che chiede di considerare la persona del paziente nella sua globalità⁵⁸. Il percorso deliberativo intrapreso da medico e paziente dovrebbe comprendere tutta questa ricchezza concettuale e metodologica ed è in virtù di ciò che il paziente potrebbe valutare se il medico è degno di fiducia oppure non lo è. Per garantire la fiducia, non basta divulgare informazioni, ma renderle intelligibili, accessibili e valutabili da tutte le parti interessate⁵⁹. La competenza epistemica del medico è strettamente connessa alla sua competenza etica, perché attraverso la comunicazione con il paziente egli rende le informazioni fruibili per ogni specifico contesto biografico oltre che clinico. Tale intreccio di competenze fa sì che il medico possa sempre fornire risposte a ogni domanda avanzata dai pazienti⁶⁰.

Nella relazione terapeutica che include i sistemi di IA possiamo ritrovare, da un lato, un medico che agirebbe come un esperto grazie alla sua formazione ed esperienza, d'altro lato, invece, un sistema di IA che agisce come un esperto medico in virtù della mole di dati disponibili. Da un lato capacità, pur fallibili, epistemiche ed etiche, d'altro lato esclusive, ma talvolta opache, capacità epistemiche. A questo punto, occorre chiedersi quanto peso il medico dovrebbe assegnare alla diagnosi formulata dall'algoritmo, soprattutto nella misura in cui il suo parere dovesse divergere dal prodotto della macchina. Il medico dovrebbe decidere se mantenere la sua convinzione oppure mutarla⁶¹. In quest'ultimo caso, però, egli perderebbe la sua autorità epistemica rispetto alla decisione clinica assunta, ma anche la sua autorità etica, in quanto non sarebbe in grado di dialogare in maniera trasparente con il paziente e adeguata alla peculiarità di quest'ultimo.

58 J.A. Marcum, *An Introductory Philosophy of Medicine: Humanizing Modern Medicine*, Springer, Dordrecht, 2008.

59 G. Starke, *The Emperors's New Clothes? Transparency and Trust in Machine Learning for Clinical Neuroscience*, in Friederich et alii, (eds.), *Clinical Neurotechnology meets Artificial Intelligence, Advances in Neuroethics Cham*, Springer, 2021, pp. 183- 196; L. Arbelaex Ossa, G. Starke, G. Lorenzini, J.E. Vogt, D.M. Shaw, B.S. Elger, *Re-focusing explainability in medicine*, in *"Digital Health"*, 8, 2022; W.K. Diprose, N. Buist, N. Hua, Q. Thurier, G. Shand, R. Robinson, *Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator*, in *"Journal of the American Medical Informatics Association"*, 27(4), 2022, pp. 592-600.

60 A. Goldman, *Expertise*, in *"Topoi"*, 37, 2018, pp. 3-10; Id., *Experts: Which Ones Should You Trust?*, in *"Philosophy and Phenomenological Research"*, 63, 2001, pp. 85-110.

61 T. Grote, P. Berens, *On the ethics of algorithmic decision-making in healthcare*, in *"Journal of Medical Ethics"*, 46, 2020, pp. 205-211.

Se è vero che gli sviluppi dell'IA si stanno rivelando uno strumento fondamentale in medicina, è tuttavia vero anche che essi stanno mettendo in crisi la possibilità di un processo decisionale condiviso all'interno della relazione terapeutica. Il tramonto dell'autorità epistemica ed etica del medico a favore di un efficiente ma ancora opaco sistema operativo potrebbe mettere a repentaglio l'autonomia decisionale del paziente, che si radica nella possibilità che il paziente ha di fidarsi del medico.

A parere di chi scrive la problematicità non interessa l'eventuale estinzione della figura del medico, ma, dal momento che i sistemi di IA potrebbero assumere un notevole margine di autonomia decisionale, ciò potrebbe stravolgere gli assetti attuali di responsabilità, dando forma ad azioni imprevedibili e inspiegabili. Ne risulterebbe una forte lacuna nella tutela del paziente, che non avrebbe più gli strumenti adeguati per valutare le prestazioni della macchina e garantirsi da eventuali pratiche indesiderate o addirittura dannose. I medici dovrebbero mantenere il controllo dei sistemi sanitari e delle decisioni mediche, mentre le informazioni e il sostegno del medico dovrebbero essere facilmente accessibili e consultabili al paziente, al fine di un corretto uso delle tecnologie.

Questo non vuol dire, dunque, rinunciare ai benefici che senza dubbio l'IA può apportare in medicina, ma promuovere una riflessione critica sull'autonomia del paziente e sulla pratica del consenso informato alla luce di una relazione terapeutica irreversibilmente mutata.