

# In primo luogo calcola di non far danni

Abbiamo bisogno di principi e regole nuove per governare l'intelligenza artificiale in medicina

I medici sono guidati dal giuramento di Ippocrate a non recare danno, e dobbiamo sicuramente aspettarci lo stesso dall'intelligenza artificiale applicata alla medicina. Un tempo considerata una tecnologia del futuro, l'intelligenza artificiale sta diventando sempre più comune. Con un sistema sanitario che lotta per mantenere la propria forza lavoro e gestire i costi, e di fronte alla crescente complessità del servizio, vi è una evidente spinta a delegare compiti di routine all'intelligenza artificiale.

L'intelligenza artificiale di oggi si presenta in molte forme<sup>1</sup>. Le più comuni sono gli analytic tool per scopi specifici che trovano applicazione ovunque, controllando automaticamente i risultati di laboratorio o le immagini diagnostiche, fino alla segnalazione precoce di focolai epidemici. Questi strumenti utilizzano algoritmi creati con il machine learning, dove la "intelligenza" risiede nel processo di apprendimento dai dati archiviati. Gli agenti di conversazione come i chatbot offrono risposte intelligenti adattate a un insieme più ampio di circostanze e possono ad esempio vedersi affidato il compito di triaging dei pazienti in base ai loro sintomi, come il National health system sta ora sperimentando<sup>2</sup>. L'intelligenza artificiale intesa come entità autonoma e capacità di ragionare



**Enrico Coiera**  
Direttore Centre for health informatics  
Australian institute of health innovation  
Macquarie university,  
Sydney (Australia)



**Maureen Baker**  
Chief Shadow board  
Faculty of clinical informatics  
(Gran Bretagna)



**Farah Magrabi**  
Centre for health informatics  
Australian institute of health innovation,  
Macquarie university,  
Sydney (Australia)

in senso ampio è nota come intelligenza generale forte (artificial general intelligence), e resta anni lontana.

La information technology modella fondamentali i processi decisionali umani ma, se mal progettata o utilizzata, può causare danni ai pazienti<sup>3</sup>. Notoriamente le leggi di Asimov obbligarono i robot a non fare del male o uccidere un essere umano o, attraverso l'inazione, a non permettere a un essere umano di essere danneggiato, e queste leggi dovevano essere scritte nel loro dna digitale. Eppure ogni nozione di danno e benefi-

cio è plasmata culturalmente e definita in modo individuale. Mentre le leggi di Asimov originate nella finzione hanno fornito l'impalcatura basilare per pensare alla sicurezza dell'intelligenza artificiale. Semplici regole che, sfortunatamente, diventano presto inutili nella complessa giungla etica che è l'assistenza sanitaria<sup>4</sup>.

Cosa accadrebbe ad esempio se l'intelligenza artificiale fosse ammessa al processo decisionale di fine vita? I medici si stanno già interrogando se algoritmi che valutano il rischio di morte saranno un elemento scatenante di discussioni con pazienti e familiari<sup>5</sup>. È accettabile che degli algoritmi oggi, o una intelligenza generale forte tra un decennio, suggeriscano l'interruzione di cure aggressive e quindi l'accelerazione del processo del morire? O in alternativa, se raccomandasse la prosecuzione di cure superflue? La nozione del "non nuocere" si estende ulteriormente quando un'intelligenza artificiale si trovi a dover scegliere tra beneficio per il paziente e vantaggio per la società. Abbiamo quindi bisogno di sviluppare principi di ampie vedute per governare la progettazione, la creazione e l'uso dell'intelligenza artificiale nell'assistenza sanitaria. Questi principi dovrebbero tenere in considerazione i tre domini rappresentati dalla tecnologia, dagli utenti e dal modo con cui entrambi interagiscono nel sistema sanitario (socio-tecnico)<sup>6,7</sup>.

Innanzitutto, come qualsiasi tecnologia, l'intelligenza artificiale deve essere progettata e costruita per soddisfare gli standard di sicurezza, che assicurano che sia adatta allo scopo e funzioni come previsto<sup>8</sup>, e le esigenze di coloro che lavoreranno con essa adattandosi ai loro flussi di lavoro<sup>9</sup>. Una cosa è non fare del male, altra è perdere tempo o far lavorare di più le persone perché si adeguino all'intelligenza artificiale. Il punto in cui gli esseri umani saranno completamente esclusi dal ciclo decisionale e le attività saranno delegate alla macchina varierà in base all'attività e al setting. Decidere a p.30 →

**Abbiamo quindi bisogno di sviluppare principi per governare l'intelligenza artificiale nell'assistenza sanitaria.**

genza artificiale? C'entra perché la tecnologia non si è ancora dimostrata utile per migliorare l'aderenza alle prescrizioni del medico, tanto che riguardino l'assunzione di medicinali quanto lo stile di vita. Emanuel cita diversi esempi, tra cui uno studio che ha messo a confronto pazienti con scompenso cardiaco teleassistiti con altri assegnati a un'assistenza ordinaria e che non ha mostrato differenze di sopravvivenza o di nuovi ricoveri. "Il monitoraggio continuo promesso dalla tecnologia è evidentemente uno spreco", anche perché le persone che più potrebbero beneficiarne, gli anziani, sono quelle che hanno meno dimestichezza con il tech. Gli argomenti sollevati dalla nota sul *Wall Street Journal* si ricollegano allo stato della mobile health, che non esaurisce le potenzialità del machine learning ed è solo un aspetto molto elementare dell'apporto delle tecnologie alla salute. Un articolo sul *New England Journal of Medicine* sottolinea come stia ancora muovendo i primi passi: a fronte delle oltre 1200 app mediche già catalogate, non disponiamo di prove della loro efficacia per migliorare la salute dei cittadini: "la gran parte dei programmi non è valutata in modo sensato e la letteratura è dominata da casi clinici e studi di piccole dimensioni"<sup>10</sup>.

Queste critiche sembrano a molti un po' conservatrici: davvero siamo fermi alla contrapposizione tra high-tech e high-touch, tra tecnologia ed empatia e contatto umano?

Rebecca De Fiore, Corso Reporting,  
Scuola Holden, Torino

1. Columbus L. Gartner's Hype cycle for emerging technologies, 2017 adds 5G and deep learning for first time. Forbes, 15 agosto 2017.
2. Feun J, Raskino M. Mastering the hype cycle. Boston: Harvard Business Press, 2002.
3. Oremus W. The never-ending hype cycle. Slate, 4 giugno 2014.
4. Chen JH, Asch SM. Machine learning and prediction in medicine: beyond the peak of inflated expectations. *New Engl J Med* 2017;376:2507-9.
5. Hixon T. Digital health slides down the back side of the hype curve. Forbes, 10 ottobre 2016.
6. Ambrosino F. Saper guardare dietro le apparenze. *Recenti Prog Med* 2017; suppl Forward 5; S29-S31.
7. Emanuel EJ. The hype of virtual medicine. *Wall Street J*, 10 novembre 2017.
8. Roess A. The promise, growth, and reality of mobile health: another data-free zone. *New Engl J Med* 2017;377:2010-1.





da p.29 → quando verrà raggiunto quel punto richiederà metodi che dovranno valutare la sicurezza, l'efficacia e l'accettabilità. Tali test non potranno del tutto garantire la sicurezza dei comportamenti di un'intelligenza artificiale complessa nel real world, pertanto saranno necessari un monitoraggio prospettico e procedure a prova di guasti (*fail-safe*). Le intelligenze generali forti future potrebbero anche avere una consapevolezza intrinseca dei principi di sicurezza per consentire loro di navigare in situazioni inaspettate<sup>10</sup>.

In seconda istanza sono necessari dei principi che garantiscano l'attendibilità delle deduzioni o le azioni di un'intelligenza artificiale. Forse, questa dovrebbe essere in grado di spiegare come è giunta a una conclusione e fornire prove a sostegno del proprio ragionamento<sup>11</sup>. Un corollario potrebbe essere che gli umani abbiano il diritto di contestare la decisione di una intelligenza artificiale se credono che sia caduta in errore. La spiegazione è semplice quando la conoscenza all'interno di un'intelligenza artificiale è esplicita, come una regola clinica, nel senso che può indicare l'applicabilità della regola alla situazione attuale. Tuttavia, per le intelligenze artificiali basate su reti neurali di oggi dare una spiegazione è difficile perché la conoscenza non è più esplicita ma codificata in modo non trasparente nelle connessioni tra "neuroni".

Altri problemi emergono quando le intelligenze artificiali sono costruite usando il machine learning, estrapolando la conoscenza da grandi insiemi di dati. Se i dati sono inaccurati o mancanti, potrebbero non essere adatti allo scopo di informare un'intelligenza artificiale o di prendere decisioni. Se i dati sono distorti, lo sarà anche la conoscenza dell'intelligenza artificiale. I bias nascosti possono discriminare alcuni pazienti in base al genere, all'etnia o alla malattia, perché sottorappresentati nei dati originali utilizzati per "formare" l'intelligenza artificiale<sup>12</sup>.

Abbiamo anche bisogno di principi che governino il modo di usare l'intelligenza artificiale che non dovrebbe essere spinta oltre i limiti dell'obiettivo della sua progettazione o dell'autorità che gli è stata delegata. Dovremmo riconoscere che le nostre prestazioni sono alterate quando lavoriamo con l'intelligenza artificiale. Il cosiddetto *automation bias*,

per esempio, si verifica quando le persone delegano troppo alla tecnologia e riducono la vigilanza<sup>13</sup>. Di conseguenza, gli esseri umani possono cadere "fuori dal circuito", non notare eventi critici o non comprendere l'accaduto al punto di non riuscire a recuperare la situazione dopo un incidente. Se le persone sono responsabili di un risultato, dovrebbero essere obbligate a rimanere vigili, anche dopo aver delegato dei compiti a un'intelligenza artificiale.

Queste considerazioni vanno ben oltre lo scopo degli attuali processi regolatori per i dispositivi medici. Includono discussioni su valori ed etica, ed è probabile che richiedano frequenti revisioni man mano che crescono competenze e ruoli per l'intelligenza artificiale<sup>14</sup>. L'Unione europea sta sviluppando un framework per le implicazioni giuridiche ed etiche dei robot<sup>15</sup>. L'assistenza sanitaria deve anche impegnarsi nello sviluppo di principi e di strutture normative e di governance che rendano la transizione verso un sistema sanitario compatibile con l'intelligenza artificiale il più sicuro ed efficace possibile. La governance dei singoli sistemi di intelligenza artificiale è probabilmente di chi è già incaricato della supervisione o della regolamentazione dell'information technology nel settore sanitario. Potrebbe richiedere uno sforzo significativo l'adattamento ai diversi profili di rischio di questa tecnologia. Il governo, i servizi sanitari, i gruppi di consumatori e le professioni cliniche dovranno concentrarsi sulle implicazioni operative. Per esempio, l'intelligenza artificiale richiederà dei cambiamenti dalla formazione ai flussi di lavoro clinici.

**“L'intelligenza artificiale metterà in discussione la nostra concezione del lavoro clinico, il dovere professionale e la natura stessa e la progettazione dei servizi sanitari.”**

Se incarnata in un robot o dispersa attraverso una rete di computer, l'intelligenza artificiale metterà in discussione la nostra concezione del lavoro clinico, il dovere professionale e la natura stessa e la progettazione dei servizi sanitari. Se riusciremo a farlo nel modo giusto, creeremo un mondo in cui noi, come clinici o pazienti, lavoreremo al sicuro fianco a fianco con i nostri preziosi amici computer. ▶

Questo articolo riprende il post degli autori pubblicato su *BMJ Opinion*. Ringraziamo gli autori e il *BMJ* per l'autorizzazione.



- Coiera E. Guide to health informatics (3rd edition). London: CRC Press, 2015.
- Burgess M. The NHS is trialling an AI chatbot to answer your medical questions. *Wired*, 5 gennaio 2017.
- Kim MO, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. *J Am Med Inform Assoc* 2017;24:246-50.
- Clarke R. Asimov's laws of robotics: implications for information technology - Part I. *Computer* 1993;26:53-61.
- Cardona-Morrell M, Chapman A, Turner RM, et al. Pre-existing risk factors for in-hospital death among older patients could be used to initiate end-of-life discussions rather than rapid response system calls: a case-control study. *Resuscitation* 2016;109:76-80.
- Coiera E. Four rules for the reinvention of healthcare. *BMJ* 2004;328:1197-99.
- Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. *Qual Saf Health Care* 2010;19(Suppl 3):i68-i74.
- Fox J, Das S. Safe and sound: artificial intelligence in hazardous applications. Cambridge Mass: MIT Press, 2000.
- Coiera EW. Artificial intelligence in medicine: the challenges ahead. *J Am Med Inform Assoc* 1996;3:363-66.
- Ong MS, Magrabi F, Coiera E. Syndromic surveillance for health information system failures: a feasibility study. *J Am Med Inform Assoc* 2012;20:506-12.
- Shortliffe EH, Axline SG, Buchanan BG, et al. An artificial intelligence program to advise physicians regarding antimicrobial therapy. *Comput Biomed Res* 1973;6:544-60.
- Reynolds M. Bias test to prevent algorithmic discriminating unfairly 201. *New Scientist*, 29 marzo 2017.
- Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *J Am Med Inform Assoc* 2017;1:24:423-31.
- Anderson M, Anderson SL. Machine ethics. Cambridge: Cambridge University Press, 2011.
- Committee on legal affairs. Draft report with recommendations to the Commission on civil law rules on robotics: European Parliament, 2016.