Il problema “Black Box” nella Medicina Digitale

Data 17 dicembre 2021
Categoria Medicinadigitale

L’intelligenza artificiale è come il cervello: non si può tagliare la testa e vedere come funziona

Il concetto di “black box”, ovvero “scatola nera” è stato formulato nella Teoria dei Sistemi ancora nel secolo scorso: con questo termine si intende un sistema che sia descrivibile nel suo comportamento esterno (output) ovvero solo per come reagisce in uscita ad una determinata sollecitazione in ingresso (input), ma il cui funzionamento interno sia ignoto.
Il termine è spesso utilizzato nella letteratura sul “digitale”,inizialmente riferendosi agli algoritmi che guidano i computer, ignoti ai semplici utilizzatori, e più recentemente a tutti I dispositivi di Deep Learning ,indecifrabili non solo per gli utenti ma spesso anche per l programmatori…
In questa pillola riporteremo alcune brevi ma illuminanti osservazioni di alcuni grandi esperti di IA, e proporremo alcune riflessioni sui risvolti etici e psico-sociali di questo fenomeno, in particolare per quel che riguarda la relazione medico-paziente.

L’intelligenza artificiale è come il cervello: non si può tagliare la testa e vedere come funziona

Questa frase di Andy Rubin, cofondatore di Android , riassume con efficacia un dato importante e imbarazzante per gli esperti di intelligenza artificiale (IA), spesso non in grado di sapere come il sistema di IA “ragioni” e arrivi a proporre una scelta decisionale. Nello Cristianini, grande esperto, ha espresso bene la situazione definendo la IA come una sorta di genialità aliena: non sappiamo bene come funzioni, abbiamo rinunciato a comprendere il perché, ma non vi è dubbio che in moltissimi settori funzioni . Per questo tali procedure vengono definite “opache ma efficienti”. L’idea prevalente è addirittura che tanto più complesso è il modello tanto migliori siano le sue performance.
Nel momento in cui un modello di deep learning, il più soggetto al problema, prospetta l’indicazione ad una indagine diagnostica, per esempio ad una biopsia cutanea poiché, con elevata probabilità, ha stabilito che si tratta di un melanoma, nessuno può stabilire sulla base di quali caratteristiche della lesione la macchina abbia elaborato questa predizione, tanto che la modalità operativa di questi sistemi è stata definita come modello black box, ovvero scatola nera . In genere tale inesplicabilità dipende dalla complessità di elaborazione dei dati, non alla portata dell’intelligenza umana, talvolta dalla segretezza per la proprietà del brevetto.
Secondo Eric Topol, l’aspetto black box dell’IA sarebbe peraltro enfatizzato a causa delle eccessive attese. Gli algoritmi non sono infallibili e non sono trasparenti nei loro passaggi computazionali, ma anche molti aspetti della pratica clinica sono poco chiari (ad esempio la prescrizione di terapie delle quali non si conosce completamente il meccanismo di azione), ma sono in generale maggiormente tollerati, in quanto “umani”.
Un approccio per sollecitare i medici ad accettare questo sistema “imperscrutabile” potrebbe essere di utilizzare quanto prodotto dal software di apprendimento per addestrare un altro modello, più trasparente, che fornisca risposte analizzabili e comprensibili degli umani . Peraltro, secondo alcuni esperti, i modelli esplicabili dovrebbero essere evitati, mentre, in alcuni ambiti, per esempio nella giustizia, nell’assistenza sanitaria e nella computer vision, potrebbero sostituire quelli a scatola nera .

La responsabilità delle scelte
L’utilizzazione della IA nelle procedure diagnostiche e terapeutiche muta radicalmente il ruolo del medico e la sua relazione con il paziente: l’elemento black box-scatola nera è un’ulteriore fattore di complicazione perché non consente al medico di conoscere i passaggi che hanno condotto la macchina alla decisione e questo determina difficoltà nella condivisione delle scelte con il paziente:”non so bene perché ma il computer dice che devi operarti”.
L’ attribuzione delle scelte decisionali dovrebbe rimanere prerogativa del medico, pur condivise a priori con il paziente, sia nel caso che il professionista decida di avvalersi dei sistemi di IA, sia che decida di non avvalersene. In generale esiste peraltro la possibilità che si sviluppi un meccanismo psicologico di de-responsabilizzazione e di delega dell’intelligenza naturale a quella artificiale: “non sono io che ho sbagliato”, con relative conseguenze legali tutte da considerare in un ambito ancora sostanzialmente sconosciuto.
Tutto può procedere bene finchè non si verificano errori e danni a carico del paziente. Se l'errore è palesemente del medico la situazione, per quanto spiacevole, è tuttavia chiara: ne risponde il medico a tutti gli effetti. Ma se l'errore è compiuto dal sistema di IA, per giunta in maniera imperscrutabile, chi ne risponde? Il problema è aperto e molto complesso.
La Commissione Europea ha pubblicato recentemente un documento che esamina le implicazioni legali dell’IA ad ampio raggio. Il documento esamina tra l’altro il tema della responsabilità per gli eventuali danni . La conclusione, in estrema sintesi, è che restano molte lacune che richiedono correzioni delle norme attualmente in uso. In mancanza di contributi chiari ed esaurienti su questo delicato tema, si ritiene utile proporre un approccio metodologico che possa fornire ragionevoli garanzie tanto al paziente quanto al medico .

1) Il cittadino-paziente deve essere informato che il medico si avvale di uno o più sistemi di intelligenza artificiale e deve autorizzare il medico a procedere nella sua utilizzazione.

2) Nella informazione al paziente il medico dovrebbe esplicitare chiaramente le probabilità di errore dei sistemi di intelligenza artificiale, sulla base dei dati disponibili per quei sistemi e per quella categoria di problemi sanitari.
3) Se il paziente acconsente alla utilizzazione della IA, il medico dovrebbe registrare in percorsi differenziati le decisioni e le conclusioni del sistema di intelligenza artificiale e quelle del proprio processo analitico, chiarendo su quali basi ritenga di accettare o rifiutare le conclusioni della intelligenza artificiale.
4) In caso di errore del medico si procede come di consueto; nel caso di errore del sistema di intelligenza artificiale saranno i vari esperti a stabilire a quale componente tecnologica e quindi a quale componente umana possa essere attribuita la responsabilità.
5) Se il paziente rifiuta l'apporto della intelligenza artificiale il medico utilizzerà le proprie conoscenze e capacità per risolvere il problema avendo cura di rilasciare documentazione dalla quale si possa evincere che ha agito con diligenza, prudenza e perizia.

Conclusioni
Le rivoluzioni tecnologiche hanno da sempre un ruolo creativo e distruttivo al tempo stesso. Quella determinata dalla IA ha aperto nuovi affascinanti orizzonti, in fondo ai quali si intravedono anche potenziali problemi. Sicuramente è necessario un approfondimento culturale generale per iniziare un percorso di confronto interdisciplinare. Sono indispensabili strategie e politiche rispetto alla gestione di una tecnologia che, attualmente impiegata limitatamente rispetto alle sue potenzialità, in un futuro non lontano è destinata a cambiare l'essenza della medicina e della relazione medico-paziente.

Giampaolo Collecchia e Riccardo De Gobbi

Bibliografia

Deluzarche C, Deep learning, le grand trou noir de l'intelligence artificielle, Maddyness, 2017 https://www.maddyness.com/2019/08/20/ia-deep-learning-trou-noir-intelligence-artificielle/
Cristianini Nello in: New Scientis Macchine che pensano. Dedalo Edizioni Bari 2018
Salovey P, Mayer J: Emotional Intelligence. Imagination, Cognition and Personality 1990 https://doi.org/10.2190/DUGG-P24E-52WK-6CDG
New Scientist. Macchine che pensano. La nuova era dell'intelligenza artificiale. Edizioni DEDALO, Bari, 2018
Rudin C. Stop explaining blak box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 2019; 1: 206-15
Holder C et al. Legal and regulatory implications of Artificial Intelligence. The case of autonomous vehicles, m-health and data mining. Luxembourg: Publications Office of the European Union, 2019
Collecchia G, De Gobbi R. Intelligenza artificiale e medicina digitale. Una guida critica. Roma: Il Pensiero Scientifico Editore,2020

Perapprofondimenti:
Giampaolo Collecchia e Riccardo De Gobbi: Intelligenza Artificiale e Medicina Digitale Il Pensiero Scientifico Ed. Roma 2020