Intelligenza artificiale: un uso responsabile ed etico

Ivana Bartoletti
Wipro Technologies e Privacy & Cybersecurity Fellow, Virginia Tech University

Il dibattito sull'AI Act è finalmente entrato nel vivo. È un dibattito importante per tutta la società, in particolare in ambito medico, dove sono emersi i benefici e le promesse più interessanti rispetto ai progressi dell'intelligenza artificiale (IA).

Non c'è dubbio che l'IA in ambito medico offra opportunità notevoli, inclusa la capacità di diagnosticare potenzialmente una malattia in uno stadio molto più precoce. Si pensi alla possibilità di esaminare una quantità enorme di immagini per scoprire deviazioni o dettagli altrimenti invisibili all'occhio umano; oppure di sviluppare nuove medicine, o politiche innovative di sanità pubblica, grazie a un supporto predittivo basato su dati storici e territoriali, dunque in grado di supportare una allocazione delle risorse più efficiente.
Tuttavia, non è questa la sede per valutare i benefici dell'IA nel campo tossicologico, dove la ricerca avanza rapidamente. L'obiettivo di questo contributo è presentare una disamina dei rischi connessi all'uso acritico dell'Intelligenza Artificiale e dei *Big Data*, così da cogliere i benefici senza incorrere in pericoli.

La non neutralità del dato

Il dibattito sugli elementi etici legati all'intelligenza artificiale procede in maniera confusa, come è normale che sia nel caso di tecnologie che, a mano a mano, escono dall'ambito della teoria ed entrano nel mondo reale e dalle conseguenze tangibili.

È successo qualche anno fa nel Regno Unito, con gli esami per gli A-level, che a causa della pandemia vennero cancellati: i voti finali furono quindi surrogati da un algoritmo che poi si scoprì discriminare gli studenti delle scuole statali. Più recentemente, è stato scoperto che l'algoritmo usato da Deliveroo per assegnare i turni di lavoro discriminava a sfavore di chi partecipa ad attività sindacali.
E ancora più recentemente, in un caso che è diventato un monito per l'Europa intera,

una indagine a posteriori ha rilevato che l'algoritmo Syri, usato in Olanda per assegnare sussidi sociali, ha discriminato i cittadini con doppia nazionalità: ben 20.000 famiglie sono state erroneamente private del supporto sociale, con conseguenze disastrose, specie sui più piccoli.
Da una parte siamo bombardati da messaggi che celebrano la potenza sovrumana dei dati – spesso definiti come "petrolio" – la cui analisi produrrebbe risultati obiettivi e quindi dovrebbe sostituirsi alla politica, alle decisioni di sanità o finanza.
D'altro canto, le applicazioni pratiche ci pongono davanti ad una realtà assai diversa: i dati ingeriti acriticamente da un algoritmo, lungi dall'essere obiettivi, rischiano semplicemente di automatizzare e ottimizzare le disuguaglianze attuali, che ovviamente sono riflesse nei dati stessi.

Le origini del *bias*

La manipolazione dei dati tramite algoritmi è oramai realtà quotidiana, ad esempio per il credito, l'*advertisement* personalizzato, l'accesso ai servizi e i sistemi predittivi. La società algoritmica è un modello in cui gli algoritmi hanno una funzione *editoriale* (curando la panoramica di notizie a cui abbiamo accesso), una funzione *creativa* (creando, nel caso degli algoritmi predittivi, i nessi tra passato, presente e futuro) e una funzione di *ingresso* (regolando l'accesso a servizi ed opportunità).

Questi ruoli sono costruiti sul processo di "datificazione" della nostra società, inteso nel senso di trasporre in dati calcolabili tutti i fenomeni che ci circondano, per poi estrarre conclusioni, correlazioni e, talora, causalità.
Potendo già osservare i grandi benefici del progresso tecnologico abilitato dall'IA, siamo d'altronde anche in grado di percepire, e ancor meglio analizzare, i rischi che lo accompagnano.
Dal concedere un credito maggiore all'uomo rispetto che alla donna nella stessa coppia, alla predizione di risultati scolastici migliori per studenti di scuole private rispetto a studenti di scuole statali, la softwarizzazione della discriminazione avviene in maniera oscura e schiva, comportando un rischio rappresentazionale.
I dati utilizzati per il *training* degli algoritmi, per niente neutri, rappresentano la realtà e solidificano le disuguaglianze esistenti. Un algoritmo, privo di un'adeguata riflessione sulla *fairness* e delle appropriate azioni correttive, finirà necessariamente per essere *biased*.
Nel 2019, uno studio rilevò pregiudizi razziali, con molto clamore, in un algoritmo clinico utilizzato da molti ospedali per allocare le cure ai pazienti: la soglia di malessere necessaria per l'erogazione di una stessa cura era più alta per i pazienti neri rispetto a quelli bianchi.

Fin qui

Come è potuto accadere? L'algoritmo era stato addestrato usando i dati sulla **spesa sanitaria passata**; le disparità tra pazienti neri e pazienti bianchi in termini di reddito e di ricchezza di lunga data, presenti ovviamente nei dati storici, hanno così potuto condizionare anche le scelte future.
Sebbene il pregiudizio di questo algoritmo sia stato fortunatamente rilevato e poi corretto, questo incidente solleva un dubbio importante: quanti altri strumenti clinici e medici sono affetti da simili discriminazioni?
Uno studio recente ha mostrato che uno strumento di intelligenza artificiale addestrato su immagini mediche, come i raggi X e le scansioni TC, aveva inaspettatamente imparato a distinguere la razza dichiarata dai pazienti, nonostante fosse stato addestrato solo allo scopo di aiutare i medici a diagnosticare le immagini dei pazienti.
La capacità di questa tecnologia di riconoscere la razza dei pazienti, anche quando il medico non è in grado di farlo, può quindi causare abusi e discriminazioni che sono

particolarmente difficili da rilevare e correggere.

Un uso consapevole e responsabile

Specialmente in campo medico, è quindi necessario un uso consapevole e responsabile della IA.

Nessuno dubita che queste tecnologie possano essere notevolmente importanti e di grande sostegno agli operatori medici e sanitari, a patto di rispettare la legge (la normativa della *privacy*, con il suo accento sulla trasparenza, la centralità dell'utente, i suoi diritti informativi, cioè la piena conoscenza di quanto accade alle sue informazioni), ma anche inquadrandosi in una cornice etica e di rispetto dei diritti umani. La normativa europea, attualmente in corso di elaborazione, richiederà che certe applicazioni di IA ad alto rischio siano conformi ad una serie di criteri enumerati nella legge. Probabilmente l'aspetto normativo non sarà sufficiente per costruire la IA in modo responsabile; non c'è dubbio che proprio su questo terreno dobbiamo crescere, e in fretta, in un contesto globale competitivo. Spetta a tutti fare in modo che quella crescita non sacrifichi i diritti fondamentali di persone, pazienti e cittadini.