

# Un algoritmo condanna i criminali

L'intelligenza artificiale (IA) ha ormai influenzato quasi ogni aspetto della nostra vita. Naturalmente, le previsioni delle macchine non possono sempre essere accurate al 100%. Ma il costo dell'errore non comporta nessun sacrificio. Questo quasi sempre.

Infatti c'è qualche eccezione.

Oggi usiamo molta analisi predittiva. Per esempio, quando Netflix ci raccomanda dei film che potrebbero piacerci, quando Amazon ci consiglia dei prodotti in base ai nostri gusti, o cose del genere.

Ma cosa pensereste se la tecnologia ci dovesse consigliare dei nuovi trattamenti contro il cancro o dovrebbe assegnarci dei punteggi per valutare la probabilità di essere recidivi alla giustizia?

Quando parliamo di polizia predittiva, molti pensano subito al film di fantascienza *Minority Report*, con Tom Cruise. Basato su un racconto di Philip K. Dick, il film racconta di un futuro in cui i crimini possono essere prevenuti. Questo può sembrare uno scenario inverosimile.

Tuttavia, la giustizia predittiva esiste già oggi.

Costruito su sistemi avanzati di apprendimento automatico, c'è un'ondata di nuove aziende che forniscono servizi di previsione ai tribunali; per esempio, sotto forma di algoritmi di valutazione del rischio, che stimano la probabilità di recidiva per i criminali.

Quindi, le macchine possono capire chi sarà un criminale?

Dopo il suo arresto nel 2013, Eric Loomis è stato condannato a

sei anni di carcere basandosi in parte su una previsione che avrebbe commesso altri crimini. Equivant, la società dietro il software utilizzato nel caso di Eric Loomis, afferma di aver fornito una visione a 360 gradi dell'imputato, al fine di fornire assistenza algoritmica dettagliata nel processo decisionale giudiziario.

Questa è una delle tante compagnie nel campo della giustizia predittiva che oggi vengono utilizzate negli Stati Uniti. Un recente rapporto del Centro Informazioni sulla Privacy Elettronica, rileva che gli algoritmi sono sempre più utilizzati in tribunale per impostare la libertà su cauzione, determinare le sentenze e persino contribuire a determinare la colpevolezza o l'innocenza.

Sembra incredibile, ma è così.

Tuttavia, un rapporto investigativo di ProPublica ha rilevato che questi algoritmi tendono a rafforzare il pregiudizio razziale delle forze dell'ordine. Infatti le valutazioni algoritmiche tendono a dichiarare gli imputati neri, come criminali futuri a un tasso quasi doppio di quelli dei bianchi. Inoltre, i giudici che hanno fatto affidamento su queste nuove tecnologie, in genere non hanno capito come sono stati calcolati i punteggi.

Ora qual è il problema?

Questi sistemi di IA sono affidabili se i dati immessi sono corretti e rilevati con metodi scientifici. Se i dati sottostanti sono distorti in qualsiasi forma, c'è il rischio che disuguaglianze strutturali e pregiudizi non siano solo replicati, ma anche amplificati.

Persino piccole irregolarità possono produrre una differenza enorme. Ne sanno qualcosa chi affronta il problema della metodologia della ricerca.

Facciamo un esempio.

Google ha fatto un recente esperimento chiamato "Quick, Draw!" con cui invitava gli utenti internet di tutto il mondo a partecipare a un divertente gioco di disegno. In ogni round del gioco, gli utenti sono stati sfidati a disegnare un oggetto in meno di 20 secondi. Più di 20 milioni di persone provenienti da 100 nazioni hanno partecipato al gioco, con oltre 2 miliardi di disegni diversi di tutti i tipi di oggetti, inclusi gatti, sedie, cartoline, farfalle, skyline, ecc.

Ma quando i ricercatori hanno esaminato i disegni delle scarpe nel set di dati, si sono resi conto che un gran numero di utenti aveva disegnato scarpe che sembravano Converse. Senza volerlo si sono resi conto che il sistema di intelligenza artificiale aveva imparato a cogliere i tipici attributi visivi delle scarpe da ginnastica come l'esempio "tipo" di scarpa. Di conseguenza, le scarpe che non sembravano scarpe da ginnastica, come quelle con tacchi alti, ballerine o zoccoli, non erano riconosciute come scarpe.

Questo è un semplicissimo esempio di come il libero fruire dei dati possa comportare delle storture.

Recenti studi dimostrano che, se non vengono controllati, i modelli di apprendimento automatico apprenderanno stereotipi di genere obsoleti, come il fatto che i "medici" sono di sesso maschile e le "receptionist" di sesso femminile. In modo simile, i modelli di IA, addestrati su immagini di ex presidenti degli Stati Uniti, hanno previsto come probabili nuovi presidenti solo i candidati maschi.

Allora come possiamo progettare una IA equa?

Nell'ottobre 2018, la Conferenza internazionale dei responsabili della protezione dei dati e della privacy ha pubblicato la Dichiarazione sull'etica e la protezione nell'intelligenza artificiale. La dichiarazione afferma che "i pregiudizi illegittimi o le discriminazioni che possono

derivare dall'uso dei dati nell'intelligenza artificiale dovrebbero essere ridotti ed eliminati".

Questo potrebbe essere fatto, ma l'IA deve essere valutata su un insieme più ampio di criteri etici e legali; non solo sulla base dell'accuratezza della classificazione e delle matrici.

É ovvio che oggi ci sono dei problemi che devono essere risolti. Per esempio tutti i soggetti dovrebbero avere pari possibilità di essere rappresentati nei dati, oppure i sistemi di apprendimento automatico dovrebbero evitare effetti ingiusti sugli individui, in particolare gli impatti relativi a vulnerabilità sociali e fisiche e altri attributi sensibili, come razza, etnia, genere, nazionalità, orientamento sessuale, religione e credo politico.

É ovvio che anche l'inclusione di queste variabili non risolve il problema, ma la speranza è che, in futuro, questo approccio possa aiutare a costruire un sistema di intelligenza artificiale "neutrale" che possa davvero aiutarci.

Ma se anche l'IA ha gli stessi pregiudizi e BIAS della mente umana, questo non porta lontano, anzi non fa altro che rafforzare le nostre mancanze.