

IA, tutti temono la fine dell'umanità. Intanto gli algoritmi decidono chi viene curato, assunto o arrestato

2026-09-17 16:30:03 di Mary Whitfill Roeloffs

URL:<https://redazione.forbes.it/2026/09/17/ai-pregiudizi-rischi-concreti-algoritmi/>

Le richieste di cambiamenti politici e gli avvertimenti riguardo a una possibile presa di controllo dei sistemi essenziali, tale da portare all'estinzione, hanno portato i timori dell'opinione pubblica riguardo **a un “giorno del giudizio” causato dall'intelligenza artificiale** ai massimi livelli di sempre. Ma gli esperti sostengono che le comunità emarginate stiano già subendo i pregiudizi dell'IA, che possono negare l'accesso all'assistenza sanitaria, influenzare l'assegnazione degli alloggi o automatizzare le sentenze penali.

Fatti chiave:

- **Joy Buolamwini, fondatrice dell'Algorithmic Justice League**, sostiene che incentrare i rischi dell'IA sulle “minacce esistenziali” sia intrinsecamente elitario e distolga l'attenzione dai danni immediati già inflitti alle comunità più povere.
- La ricerca di Buolamwini si concentra su ciò che lei definisce **lo “sguardo codificato”**, un concetto ormai dimostrato secondo cui i pregiudizi e le preferenze dei creatori di un sistema vengono codificati direttamente nel software e, poiché i set di dati utilizzati per addestrare l'IA sono stati storicamente sbilanciati a favore degli uomini bianchi, lo sono anche i modelli stessi.
- Uno [studio intitolato Gender Shades](#) ha rilevato che i principali strumenti di riconoscimento facciale presentavano tassi di errore inferiori all'1% per gli uomini dalla pelle chiara, ma fino al 35–47% per le donne dalla pelle scura. Questi errori tecnici hanno portato ad **arresti ingiustificati di uomini neri** in tutto il Paese.
- È emerso che gli algoritmi utilizzati dalle società di gestione immobiliare per valutare i potenziali inquilini respingono candidati qualificati a basso reddito **senza fornire spiegazioni** né possibilità di ricorso.
- Una società di selezione degli inquilini, SafeRent Solutions, è stata costretta a risolvere una causa legale con un risarcimento di 2,3 milioni di dollari dopo che oltre 400 ricorrenti hanno sostenuto che il suo algoritmo assegnasse in modo sproporzionato punteggi più bassi ai candidati e agli inquilini afroamericani e ispanici che utilizzavano buoni abitativi governativi.
- [Uno studio di riferimento pubblicato su Science](#) ha rivelato che un algoritmo utilizzato nel settore sanitario commerciale ha **dato la priorità ai pazienti bianchi nell'assegnazione di cure mediche aggiuntive rispetto ai pazienti neri** che presentavano lo stesso grado di malattia, poiché utilizzava la spesa sanitaria passata come indicatore delle esigenze sanitarie. Poiché storicamente ai pazienti neri vengono destinati meno fondi a causa delle disuguaglianze sistemiche, il sistema ha concluso che i pazienti neri fossero “più sani”.
- Sono stati riscontrati pregiudizi algoritmici anche [nei processi automatizzati di selezione](#) e assunzione del personale, nella [polizia predittiva](#) e nelle [valutazioni dei rischi](#) relative al benessere dei minori.

Dichiarazione cruciale

“L'intelligenza artificiale è brava a descrivere il mondo così com'è oggi, con tutti i suoi pregiudizi, ma non sa

come dovrebbe essere il mondo”, ha affermato Joanne Chen, partner della società di investimenti nel settore dell’IA Foundation Capital, in occasione del SXSW del 2018.

Fatto sorprendente

Amazon [ha abbandonato](#) un motore interno di reclutamento basato sull’intelligenza artificiale dopo aver constatato che questo sistematicamente **declassava le candidature che includevano la parola “femminile”** e quelle provenienti da atenei esclusivamente femminili. L’intelligenza artificiale aveva appreso autonomamente che i candidati maschi fossero preferiti, poiché era stata addestrata su 10 anni di dati storici relativi alle assunzioni provenienti da un settore tecnologico dominato dagli uomini.

Gli allarmi lanciati dai ricercatori

Gli amministratori delegati delle principali aziende del settore dell’intelligenza artificiale chiedono **un rallentamento nello sviluppo dei modelli all’avanguardia** per consentire alle autorità di regolamentazione in materia di sicurezza e legislazione di mettersi al passo, dopo una serie di allarmi preoccupanti lanciati da esperti del settore. [Evan Hubinger](#), responsabile della scienza dell’allineamento presso Anthropic, la scorsa settimana ha affermato di ritenere che vi sia una probabilità superiore al 10% che l’umanità si estingua a causa di un’IA non allineata entro i prossimi 10 anni, mentre [Jacob Coxon](#), ex ricercatore di Anthropic e OpenAI, ha dichiarato che gli sviluppatori che realizzano questi strumenti **“credono sinceramente che potrebbe ucciderci tutti entro la fine del decennio”**. Anche alcuni altri operatori del settore hanno lanciato avvertimenti simili e oltre 1.300 dipendenti di aziende all’avanguardia nel campo dell’IA hanno aderito a un’iniziativa denominata [Pacing the Frontier](#), volta a consentire alla comprensione degli sviluppatori di stare al passo con il software che sta seguendo un percorso di auto-miglioramento.

I danni concreti che l’IA sta già facendo

C’è chi sostiene che le crescenti preoccupazioni riguardo a un “giorno del giudizio” causato dall’IA siano un prevedibile esempio di “panico tecnologico” ciclico, mentre altri ritengono che servano a distrarre l’opinione pubblica e i legislatori **dai danni concreti che l’IA sta già causando**. [Timnit Gebru](#), un’informatica che ha lasciato Google dopo aver reso pubblici i pericoli dell’IA, ha affermato che una previsione apocalittica mette in secondo piano le minacce “veramente esistenziali” dell’IA, tra cui i [licenziamenti di massa](#) attribuiti a questa tecnologia e una catastrofe climatica già in atto, aggravata dai giganteschi data center necessari per alimentare l’IA. Margaret Mitchell, anch’essa ex dipendente di Google, ha affermato che concentrarsi su scenari di minaccia esistenziale spinge i legislatori a redigere nuove normative per il settore e [li distoglie dall’applicare le leggi esistenti](#) contro **lo sfruttamento del lavoro, la violazione del diritto d’autore, il degrado ambientale e la parzialità algoritmica**. Altri sostengono che la società abbia una lunga storia di considerare gli strumenti trasformativi come minacce esistenziali; [il miliardario Marc Andreessen](#) ha paragonato l’IA alle evoluzioni tecnologiche del passato — come il fuoco — e sostiene che essa amplificherà la produttività piuttosto che sostituire l’azione umana.