



I bias di genere e razziali nell'AI

Quante volte abbiamo sentito dire che **le intelligenze artificiali (AI) sono razziste o sessiste**? Pensiamo come la selezione automatica dei curriculum vitae può influenzare la scelta dei recruiter nell'assumere donne per determinati ruoli. Oppure la rappresentazione stereotipata delle donne nei videogiochi. Come si genera un pregiudizio? Ognuno di noi, nella vita di tutti i giorni, è influenzato da pregiudizi.

Vi riportiamo questo semplice indovinello che può facilmente dimostrarlo:

Un giorno un uomo ha un incidente d'auto e in macchina con sé ha il figlio.
Il padre muore, mentre il figlio viene portato d'urgenza in ospedale.
Il chirurgo che dovrebbe operarlo, però, appena lo vede esclama: «Non posso operarlo. È mio figlio!»

Com'è possibile? Soluzione: il chirurgo è la madre del bambino.

Molti di noi hanno sicuramente avuto difficoltà nell'indovinare la risposta corretta, probabilmente perché l'utilizzo della parola chirurgo al maschile tende a indurci verso pregiudizi. Questo comportamento diffuso si riflette anche nell'intelligenza artificiale. Ma sono le AI sessiste o c'è altro dietro?

Quando parliamo di Intelligenza artificiale dobbiamo sempre ricordare che siamo noi ad avere un ruolo fondamentale nella costruzione di queste: siamo noi a generarle. La scrittrice **Cathy O'Neil**, nel suo libro *Armi di distruzione matematica* approfondisce il funzionamento degli algoritmi che ormai dominano la nostra quotidianità iperconnessa e che,

se addestrati in modo errato, sono spesso vere e proprie armi di distruzione matematica.

Purtroppo, questi bias non sono solo di genere ma anche etnici. Spesso e volentieri questi due problemi non vengono affrontati simultaneamente, eppure la matrice originale è la stessa. Sono due **battaglie intersezionali** che non possono essere separate. I bias razziali, infatti, hanno lo stesso funzionamento di quelli di genere: derivano principalmente dai pregiudizi che un soggetto ha nella sua mente. Nel pratico, gli algoritmi di AI apprendono da dati storici che spesso riflettono disuguaglianze sociali esistenti.

Vengono, quindi, riprodotti quelli che sono i limiti e gli errori della mente umana, e l'intelligenza artificiale non è così intelligente da distinguere ciò che è giusto e ciò che non lo è. Di seguito riportiamo alcuni casi:

- **Riconoscimento Facciale**

Studi hanno dimostrato tassi di errore più alti per individui con pelle scura in sistemi di riconoscimento facciale ampiamente utilizzati (Buolamwini & Gebru, 2018).

- **Valutazione del Credito**

Algoritmi di scoring creditizio hanno mostrato tendenze a sfavorire ingiustamente minoranze razziali (Fuster et al., 2020).

- **Sistemi di Giustizia Penale**

Strumenti di valutazione del rischio utilizzati nei tribunali hanno esibito bias razziali nelle loro previsioni (Angwin et al., 2016).

Da qui nasce l'urgente necessità di creare **modelli** che siano **eticamente responsabili** e di inserire figure professionali che ne valutino la correttezza. Attualmente esistono già dei metodi, tra cui gli "audit algoritmici" o il "debiasing dei dati" che però non vengono implementati e utilizzati a sufficienza.

Il razzismo algoritmico e i bias di genere nell'AI rappresentano una sfida significativa che richiede un approccio multidisciplinare. È essenziale che ricercatori, sviluppatori e politici collaborino per creare sistemi di AI più equi e inclusivi. Nel frattempo, ognuno di noi può impegnarsi attivamente per evitare di alimentare questi pregiudizi che portano a discriminazioni sistematiche di alcuni gruppi di persone.

A cura **Generazione Stem**