FONDAZIONE BRUNO KESSLER
MAGAZINE

Intelligenza artificiale e discriminazioni

4 Settembre 2023

In occasione del seminario internazionale FBK-ISR tenuto quest'estate, la ricercatrice Ilaria Valenzi ha illustrato il report “Bias in Algorithms – Artificial Intelligence and Discrimination” pubblicato dalla European Union Agency for Fundamental Rights. Le abbiamo fatto alcune domande per capire quando gli algoritmi generano discriminazioni e quali soluzioni si possono mettere in campo.

– **Di cosa si occupa la European Union Agency for Fundamental Rights (FRA) e qual è la specializzazione delle persone che vi lavorano?**

La FRA vuole contribuire all’instaurazione di una cultura dei diritti umani nell’Unione europea, attraverso la diffusione dei principi contenuti nella Carta dei diritti fondamentali. Ciò avviene attraverso azioni di promozione in numerosi settori, come l’integrazione dei migranti, le discriminazioni, la lotta al razzismo e alla xenofobia. Un particolare settore di azione è la protezione dei dati personali e l’Intelligenza artificiale. In questi ambiti, la FRA raccoglie e analizza dati, individua tendenze, collabora con le istituzioni dell’UE e gli Stati membri e gli organismi per le pari opportunità, fornisce pareri ai responsabili politici e promuove risposte politiche conformi ai diritti fondamentali. La FRA si avvale di competenze nel campo del diritto, delle scienze politiche e sociali, nelle scienze statistiche, oltre a specialisti della comunicazione.

– **Qual è stato l’obiettivo dello studio “Bias in Algorithms – Artificial Intelligence and Discrimination”?**

Lo studio ha esaminato l’utilizzo dell’intelligenza artificiale nel campo della polizia predittiva e nel rilevamento dei discorsi offensivi on line, per dimostrare come i pregiudizi negli algoritmi tendano a manifestarsi e ad amplificarsi nel tempo, influenzando la vita delle persone e potenzialmente creando situazioni discriminatorie. A partire dal presupposto della centralità dell’uso dell’intelligenza artificiale per le nostre società, lo studio ha voluto mostrare come funzionano gli algoritmi nella pratica, sviluppando dei modelli e valutandone l’affidabilità per i diritti, in particolar modo delle minoranze.

– **Ci può fare alcuni esempi di algoritmi che hanno generato delle discriminazioni?**

Sono numerosi i casi verificatisi nella pratica. Il report parte da uno dei più noti, quello che nel 2021 ha portato alle dimissioni del governo olandese. Si trattava del caso di un sistema algoritmico per l'assegnazione di sussidi sociali, che ha ingiustamente accusato 26mila famiglie di frode fiscale, famiglie che avevano tutte un background migratorio. Tra i più noti si pensi anche al caso Compas, un algoritmo di giustizia predittiva per la previsione del rischio di recidiva dei condannati, con evidenti discriminazioni etnico razziali ai danni della popolazione afroamericana. Ancora, sono numerosi i casi di intelligenza artificiale che ha "imparato" a discriminare secondo il genere o l'etnia per la selezione di curricula o per il riconoscimento facciale (in cui la macchina riconosce ad esempio il volto di persone bianche di sesso maschile ma non quello di persone nere di sesso femminile), o associando il colore della pelle esclusivamente ad alcune mansioni di lavoro. La letteratura scientifica è piena di casi del genere.

– **Che cosa è emerso dallo studio?**

I risultati evidenziano come alcuni termini legati a caratteristiche protette delle persone, come ad esempio la religione, l'etnia, il genere e l'orientamento sessuale, contribuiscono a classificare un testo, un post o un commento on line, come offensivo. Alcune appartenenze religiose, specie se declinate secondo il genere (come ad esempio l'uso del termine "musulmana" o "ebreo"), risultano le più segnalate. L'utilizzo nella lingua italiana di termini come "stranieri" aumenta la previsione di offensività. Tuttavia non sempre il modello algoritmico funziona: possono essere segnalati come offensivi commenti che in realtà non lo sono mentre sfuggono al controllo situazioni di offensività non riconosciuta. Ciò rischia di limitare alcuni diritti fondamentali o di non garantirne la tutela. La scarsa diversità linguistica costituisce un'aggravante.

– **Quali sono i principali nodi da risolvere e quali possono essere le soluzioni per il futuro?**

Si tratta di prestare particolare attenzione all'uso di algoritmi di apprendimento automatico e al processo decisionale automatizzato, assicurandosi che vengano effettuate valutazioni periodiche, migliorando la qualità dei dati. Da questo punto di vista, gli algoritmi devono essere allenati tenendo in considerazione le differenze di genere e le altre caratteristiche protette delle persone, come l'etnia, la religione, lo stato di salute. Sarà inoltre sempre più necessario promuovere la diversità linguistica per mitigare i pregiudizi negli algoritmi e restituire la varietà dei nostri contesti sociali e culturali.

LINK

https://magazine.fbk.eu/it/news/intelligenza-artificiale-e-discriminazioni/

TAG

- #ai
- #algoritmi
- #bias
- #IA
- #intelligenzaartificiale
- #scienzereligiose

AUTORI

- Viviana Lupi