Google, niente più pregiudizi razziali e di genere nei suoi sistemi di IA?

Google si impegna prima di luglio a prendere in considerazioni le ricerche effettuate dai suoi scienziati per eliminare i pregiudizi **razziali e di genere** dai suoi sistemi di intelligenza artificiale. Lo riporta l'agenzia *Reuters*.

A differenza del vertice di Google, i dipendenti favorevoli a un uso etico e responsabile dell'IA

Dunque la società alla fine sembra cedere alle pressioni interne dei dipendenti, favorevoli a un utilizzo dell'Intelligenza artificiale in modo etico e responsabile.

Prima di questo passo, Google ha licenziato una delle sue maggiori ricercatrici nell'intelligenza artificiale, alimentando le tensioni all'interno della divisione dopo l'uscita di **Timnit Gebru**, l'altra leader del team etico per l'IA di Mountain View. Le due colleghe avevano lavorato insieme a uno studio sui potenziali pregiudizi nel linguaggio dei sistemi di intelligenza artificiale come quello usato dal colosso big tech.



Timnit Gebru
Ora il dietrofront.

Le modifiche richieste dai legali di Google alle ricerche sull'IA

Reuters è venuta in possesso di un'email inviata tra i dipendenti di Google in cui emerge come l'ufficio legale della società avesse modificato uno dei tre documenti sull'intelligenza artificiale.

Uno degli autori della ricerca e mittente dell'email, **Nicholas Carlini**, ha definito le modifiche *"profondamente insidiose".*

"Siamo chiari", scrive nell'e-mail Carlini, *"quando noi accademici scriviamo che abbiamo una 'preoccupazione' o troviamo qualcosa di 'preoccupante' e un avvocato di Google ci chiede di cambiarlo allora interviene il Grande Fratello".*

Le modifiche richieste, secondo la sua e-mail, includevano,

tra l'altro, di cambiare la parola "*preoccupazioni*" in "*considerazioni*" e "*pericoli*" in "*rischi*". Gli avvocati richiedevano anche l'eliminazione dei riferimenti alla tecnologia di Google.

Evitare sistemi di IA con "bias"

Un altro documento del contendere, scrive **Isabella Corradini** su *Key4biz*, è **"On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?**". llustra i rischi dei modelli linguistici di grandi dimensioni basati su una quantità impressionante di dati testuali. Ancorché non pubblicato, lo scritto è circolato tra gli ambienti scientifici e l'essenza del contenuto è disponibile presso la MIT Technology Review.

Nella parte introduttiva del documento gli studiosi si chiedono "*se si è riflettuto abbastanza sui potenziali rischi associati al loro sviluppo e sulle strategie per mitigarli*".

Tra i pericoli evidenziati, risulta essere centrale quello della difficoltà di poter cogliere appieno tutte le sfumature dei linguaggi di regioni e popoli con scarso accesso a Internet e quindi con una minore impronta linguistica online, con l'evidente distorsione che la lingua considerata dai sistemi di IA potrebbe riflettere solo le pratiche di alcuni paesi.

Ma vi è di più. Dal momento che i modelli intelligenti di elaborazione del linguaggio si basano su una quantità sempre più crescente di testi in rete, c'è il concreto rischio che l'impossibilità di verificare la totale assenza di pregiudizi finisca con l'inglobare nella formazione dei modelli anche il linguaggio razzista, sessista e violento. Di fronte ad una mole di dati sempre più ampia da verificare, sono più che normali i dubbi sull'efficacia della metodologia e sulla sua pericolosità.

https://www.key4biz.it/google-niente-piu-pregiudizi-razziali-e

-di-genere-nei-suoi-sistemi-di-ia/347371/