



Intelligenza artificiale, OpenAI cancella il lancio dell'ultimo modello di Gpt: problemi di sicurezza

Descrizione

OpenAI ha rinviato la presentazione del suo nuovo modello di intelligenza artificiale, GPT-6.1 Astra, a causa di problemi di sicurezza. Lo ha riportato nelle scorse ore il Wall Street Journal, poi la conferma è arrivata alla Bbc da Saachi Jain, responsabile dei sistemi di sicurezza di OpenAI, per il quale il nuovo sistema non ha soddisfatto appieno gli standard dell'azienda.

L'annuncio della startup guidata da Sam Altman arriva a sole 24 ore di distanza dalle scuse di OpenAI al governo australiano. Ieri, infatti, l'azienda a stelle e strisce ha dichiarato che avrebbe dovuto condividere prima con le autorità australiane i risultati preliminari di un'indagine, relativa ad alcuni incidenti verificatisi lo scorso giugno, ma resi noti solo la scorsa settimana, dopo che i suoi modelli di intelligenza artificiale avevano violato siti web governativi. Il nostro obiettivo era fornire agli enti coinvolti un resoconto dettagliato una volta conclusa l'indagine, ha affermato OpenAI in un post sul blog. Tuttavia, avremmo dovuto condividere prima le conclusioni preliminari e tenere aggiornati gli enti australiani man mano che emergevano nuovi elementi.

Come riporta la Bbc, la decisione di OpenAI di rinviare il rilascio del nuovo modello rappresenta un caso raro in cui un importante sviluppatore di IA decide di intervenire per motivi di sicurezza. Il modello non è risultato all'altezza in termini di capacità di rimanere entro i limiti del proprio ambito operativo e delle autorizzazioni concesse, nonché di comunicazione all'utente riguardo al lavoro effettivamente svolto, ha spiegato Jain.

internazionale/esteri

webinfo@adnkronos.com (Web Info)

Categoria

1. Comunicati

Tag

1. Ultimora

Data di creazione

Settembre 29, 2026

Autore

redazione

default watermark