



ChatGPT: nuovi controlli parentali per sicurezza AI adolescenti

Descrizione

(Adnkronos) ChatGPT sta implementando una serie di nuove funzionalità volte a rafforzare la sicurezza e la supervisione per gli utenti più giovani. A partire da oggi, saranno disponibili controlli parentali e una nuova pagina di risorse per i genitori con l'obiettivo di guidare l'utilizzo di ChatGPT all'interno dei contesti familiari. Questi controlli consentono ai genitori di collegare il proprio account a quello del figlio adolescente e personalizzare le impostazioni per garantire un'esperienza sicura e appropriata all'età.

Questa iniziativa si inserisce in un più ampio sforzo di OpenAI per rendere ChatGPT utile per tutti e fornire alle famiglie gli strumenti necessari per supportare l'utilizzo dell'IA da parte dei propri figli. L'azienda ha precedentemente annunciato lo sviluppo di un sistema di previsione dell'età, destinato a identificare gli utenti di età inferiore ai 18 anni per applicare automaticamente impostazioni appropriate per gli adolescenti. I controlli parentali rappresentano un passo fondamentale in questa direzione, conferendo ai genitori la capacità di gestire l'esperienza dei propri figli con ChatGPT.

Lo sviluppo di queste funzionalità ha coinvolto una stretta collaborazione con esperti, gruppi di advocacy come Common Sense Media e legislatori, inclusi gli Attorneys General della California e del Delaware. OpenAI ha dichiarato di aspettarsi di affinare ed espandere questi controlli nel tempo. Robbie Torney, Senior Director, AI Programs, Common Sense Media, ha commentato: «Questi controlli parentali sono un buon punto di partenza per i genitori nella gestione dell'uso di ChatGPT da parte dei loro figli. I controlli parentali sono solo un pezzo del puzzle quando si tratta di mantenere gli adolescenti al sicuro online, tuttavia funzionano meglio se combinati con conversazioni continue sull'uso responsabile dell'IA, chiare regole familiari sulla tecnologia e coinvolgimento attivo nella comprensione di ciò che il loro adolescente sta facendo online».

Per attivare i controlli, un genitore o tutore deve inviare un invito al proprio figlio adolescente per collegare gli account. Una volta che il figlio accetta, il genitore può gestire le impostazioni del suo account da un'unica pagina di controllo. Anche gli adolescenti possono invitare un genitore a connettersi. In caso di scollegamento dell'account da parte del figlio, il genitore riceverà una notifica.

L'implementazione di salvaguardie migliorate per gli account degli adolescenti collegati. Una volta connessi, l'account del minore riceverà automaticamente protezioni aggiuntive sui contenuti, inclusa una riduzione di contenuti grafici espliciti, sfide virali, roleplay di natura sessuale, romantica o violenta, e ideali di bellezza estremi, al fine di mantenere l'esperienza appropriata all'età. Queste salvaguardie sono state elaborate attraverso un'attenta revisione di ricerche esistenti per comprendere le differenze di sviluppo uniche degli adolescenti. I genitori avranno la facoltà di disattivare questa impostazione se lo desiderano, ma gli utenti adolescenti non potranno apportare modifiche.

Sebbene questi sistemi di protezione siano un valido aiuto, non sono infallibili e possono essere aggirati se qualcuno sta intenzionalmente cercando di eluderli. OpenAI ha ribadito il proprio impegno a continuare a iterare e migliorare nel tempo, raccomandando ai genitori di dialogare con i propri figli sull'uso sano dell'IA e su come questo si configuri per la loro famiglia.

Ai genitori vengono offerte funzionalità aggiuntive per personalizzare l'esperienza dei propri figli. Da una semplice pagina di controllo, i genitori possono:

Impostare i "quiet hours" (ore di silenzio), ovvero orari specifici in cui ChatGPT non può essere utilizzato.

Disattivare la modalità vocale, rimuovendo l'opzione di utilizzare tale funzione in ChatGPT.

Disattivare la memoria, impedendo a ChatGPT di salvare e utilizzare le conversazioni precedenti.

Rimuovere la generazione di immagini, inibendo la capacità di ChatGPT di creare o modificare immagini.

Disattivare l'addestramento del modello, in modo che le conversazioni del figlio non vengano utilizzate per migliorare i modelli che alimentano ChatGPT.

Queste impostazioni sono opzionali e flessibili, consentendo alle famiglie di scegliere ciò che meglio si adatta alle loro esigenze.

Notifiche per situazioni di potenziale rischio

Consapevole che alcuni adolescenti possono rivolgersi a ChatGPT in momenti difficili, OpenAI ha sviluppato un nuovo sistema di notifica per aiutare i genitori a riconoscere se qualcosa non va. Sono state aggiunte protezioni che aiutano ChatGPT a riconoscere potenziali segnali che un adolescente potrebbe pensare di farsi del male. Se i sistemi rilevano un potenziale rischio, un piccolo team di persone appositamente addestrate esamina la situazione. In caso di segni di disagio acuto, i genitori verranno contattati tramite e-mail, messaggio di testo e notifica push sul telefono, a meno che non abbiano optato per non ricevere tali comunicazioni.

OpenAI sta collaborando con esperti di salute mentale e adolescenti per progettare questo sistema, riconoscendo che nessun sistema è perfetto e che potrebbero esserci falsi allarmi, ma ritiene che sia meglio agire e avvisare un genitore in modo che possa intervenire piuttosto che rimanere in silenzio. L'azienda sta inoltre lavorando a processi e circostanze appropriate per contattare le

forze dell'ordine o altri servizi di emergenza, ad esempio in caso di minaccia imminente per la vita e impossibilit  di raggiungere un genitore. Anche in queste rare situazioni, la privacy degli adolescenti viene presa sul serio, e verranno condivise solo le informazioni necessarie per consentire a genitori o soccorritori di proteggere la sicurezza del minore.

OpenAI ha creato una nuova pagina di risorse per i genitori che raccoglie informazioni su come funziona ChatGPT, i controlli parentali disponibili e idee su come gli adolescenti possono utilizzarlo per l'apprendimento, la creativita e la vita di tutti i giorni. Questa risorsa   un punto di partenza e sar  regolarmente aggiornata con suggerimenti, guide, consigli di esperti e spunti di conversazione per supportare i genitori man mano che l'uso dell'IA cresce nelle loro famiglie.

Nei prossimi mesi, OpenAI continuer  a sviluppare il sistema di previsione dell'et  che permetter  a ChatGPT di applicare automaticamente le impostazioni appropriate per gli adolescenti. Nelle situazioni in cui l'et  di un utente non   certa, verr  adottato un approccio pi  cauto, applicando proattivamente le impostazioni per adolescenti. Nel frattempo, i controlli parentali rappresenteranno il metodo pi  efficace per i genitori per assicurare che i propri figli siano inclusi nell'esperienza appropriata all'et . Questo   un passo avanti importante, ma il nostro lavoro non   finito. OpenAI continuer  a investire in funzionalit  che aiutino genitori e adolescenti a utilizzare ChatGPT in modo sicuro e con fiducia, condividendo i progressi.

  

tecnologia

webinfo@adnkronos.com (Web Info)

Categoria

1. Tecnologia

Tag

1. tec

Data di creazione

Settembre 30, 2025

Autore

redazione