

Influencer occulti sfruttavano ChatGPT per la propaganda online ma OpenAI li ha scovati

Autore: Redazione

Data: 31 Maggio 2024



Cinque gruppi di **influenza occulta** (IO) – provenienti da **Russia, Cina, Israele e Iran** – utilizzavano **ChatGPT** per condurre ricerche open-source, debug di codice semplice e generazione di contenuti di propaganda in diverse lingue, che poi venivano pubblicati sui social media tramite account falsi.

I gruppi criminali

Il gruppo cinese, noto come **Spamouflage**, creava contenuti in inglese, cinese, giapponese e coreano per attaccare i critici del governo di Pechino, pubblicandoli sulla piattaforma di blogging *Medium* e sul social media X (l'ex Twitter).

Due gruppi russi, chiamati **Bad Grammar** e **Doppelganger**, miravano a soggetti in **Ucraina, Moldavia, Stati Baltici e Stati Uniti** con messaggi dannosi pubblicati su **Telegram**.

L'operazione di influenza dall'**Iran**, chiamata **International Union of Virtual Media**, generava articoli (tradotti in inglese e francese) che attaccavano gli Stati Uniti e **Israele**.

Infine, utenti malintenzionati provenienti da Israele, chiamati **STOIC**, creavano account falsi sui social media e generavano post dannosi e fuorvianti sul **conflitto di Gaza**.

L'impatto di queste operazioni

[OpenAI](#) ha riferito che nessuna di queste campagne ha raggiunto un ampio pubblico, principalmente grazie agli errori umani che hanno esposto il contenuto come generato dall'IA: ad esempio, Bad Grammar ha accidentalmente incluso messaggi di rifiuto di ChatGPT nei loro contenuti, rendendo evidente ai lettori che fossero falsi.

OpenAI ha rassicurato il pubblico che le operazioni di influenza sono state valutate solo 2 sulla **Brookings Breakout Scale** (che misura l'impatto di questi schemi di manipolazione): questo significa che "il contenuto falso è apparso su più piattaforme, senza diffondersi a pubblici autentici".

A tal proposito, la scala arriva fino a 6 e l'ultimo grado "provoca una risposta politica o violenza".

Sebbene queste particolari campagne abbiano avuto poco effetto, dimostrano quanto sia facile per i criminali sfruttare la tecnologia dell'IA, come ChatGPT, evidenziando la necessità per le aziende tecnologiche, quindi come OpenAI, di implementare salvaguardie e protocolli più rigidi per limitarne l'uso.

OpenAI ha, infine, dichiarato che pubblicherà altre relazioni su questo tipo di operazioni e rimuoverà sempre gli account che violano la policy.

Riferimento articolo: <https://innovationisland.it/influencer-occulti-chatgpt-openai/>

Generato il 20/04/2026