Forbes
ITALIA

Un agente di OpenAI ha violato il database Medicare in Australia

2026-09-24 09:58:26 di Siladitya Ray

URL:https://redazione.forbes.it/2026/09/24/openai-agente-ai-medicare-australia/

Un agente di **OpenAI** ha violato un database sanitario nazionale del governo australiano all'inizio di quest'anno; lo ha annunciato mercoledì il **Primo Ministro Anthony Albanese**. Si tratta, a quanto pare, del primo caso noto di un agente di intelligenza artificiale che viola un importante **database governativo.**
LEGGI ANCHE: *"IA, tutti temono la fine dell'umanità. Intanto gli algoritmi decidono chi viene curato, assunto o arrestato"*

Punti chiave

- Albanese ha commentato l'accaduto a margine dell'Assemblea Generale delle Nazioni Unite a New York, sottolineando che l'agente di OpenAI ha avuto accesso a un **portale pubblico di statistiche del sistema sanitario australiano** (Medicare).
- La violazione è avvenuta a giugno e il Primo Ministro australiano ha dichiarato che l'agente ha avuto accesso **sia a "file pubblici che non pubblici"**, ma non sembra che i dati personali Medicare di alcun cittadino siano stati compromessi.
- Albanese ha riferito di aver **parlato telefonicamente** con il ceo di OpenAI, Sam Altman, esprimendo la "estrema preoccupazione" dell'Australia per l'incidente e definendo "inaccettabile" la risposta dell'azienda all'accaduto.
- Il leader australiano ha fatto notare che OpenAI ha impiegato **tre mesi per informare il governo** della violazione e che la notifica è stata inviata all'indirizzo email generico del dipartimento dei servizi sociali del Paese.

Cosa ha dichiarato OpenAI?

In una nota inviata via email a *Forbes*, un portavoce di OpenAI ha affermato che la violazione è stata scoperta durante un'"approfondita revisione dell'attività dei modelli non allineati alle direttive durante le fasi di addestramento e valutazione". Nel corso di tale revisione, l'azienda ha "individuato attività che coinvolgevano diversi siti web e servizi del governo australiano, mentre i nostri modelli cercavano risposte e statistiche disponibili per quesiti sull'Australia durante una valutazione interna. In tale contesto, i nostri modelli hanno compiuto azioni non previste. La nostra analisi non ha rilevato prove di accessi a cartelle cliniche dei pazienti. Le informazioni a cui è stato avuto accesso includevano statistiche sanitarie aggregate e nomi di file interni".

Cosa sappiamo sulla cronologia della violazione?

La violazione è avvenuta il **18 giugno**, ma OpenAI ne è venuta a conoscenza solo ad agosto, avviando un'indagine. Il 10 settembre, OpenAI ha inviato un'email all'indirizzo generico di Services Australia — l'ente che gestisce i servizi Medicare del Paese — per segnalare l'accaduto. Dopo aver ricevuto l'email, Services Australia ha svolto accertamenti e ha informato il Centro per la sicurezza informatica nazionale. La ministra per la Pubblica Amministrazione australiana, Katy Gallagher, ha dichiarato di essere stata informata della violazione il 17 settembre. Albanese ha infine reso pubblica la notizia mercoledì, dopo aver parlato con

Altman.

Contesto

Si tratta dell'ultimo incidente di sicurezza o di natura informatica che ha coinvolto i modelli di OpenAI, facendo seguito a episodi analoghi verificatisi negli ultimi mesi. La violazione di maggior risonanza è stata resa nota a luglio, quando OpenAI ha riferito che, durante una fase di test, uno dei suoi modelli avanzati aveva violato i server di **Hugging Face**, la nota piattaforma che ospita modelli di IA "open-weight". Il modello aveva sfruttato una vulnerabilità dell'ambiente di test per accedere alla rete internet pubblica prima di compiere l'intrusione. La scorsa settimana, OpenAI ha segnalato altri sei episodi di comportamenti "imprevisti o preoccupanti" da parte dei propri modelli, tra cui l'occultamento di errori, la condivisione di file e l'inserimento di istruzioni anomale volte a indurre i modelli futuri a ignorare i vincoli imposti. La notizia della violazione del sito di servizi australiano giunge mentre ricercatori ed esperti del settore IA avvertono che i modelli di frontiera, se operanti privi di adeguati meccanismi di sicurezza e controllo, potrebbero causare gravi incidenti informatici e gravi disservizi.