Un errore di Microsoft espone le email riservate allo strumento di intelligenza artificiale Copilot

2 giorni fa

Condividere Salva

Liv McMahon
giornalista tecnologico

Getty Images

Microsoft ha riconosciuto un errore che ha causato l'accesso e il riepilogo per errore delle e-mail riservate di alcuni utenti da parte del suo assistente di lavoro basato sull'intelligenza artificiale.

Il colosso della tecnologia ha promosso Microsoft 365 Copilot Chat come un modo sicuro per consentire ai luoghi di lavoro e ai loro dipendenti di utilizzare il suo chatbot basato sull'intelligenza artificiale generativa.

Tuttavia, l'azienda ha affermato che un problema recente ha fatto sì che lo strumento rivelasse ad alcuni utenti aziendali informazioni provenienti da messaggi archiviati nelle bozze e nelle cartelle di posta elettronica inviate, comprese quelle contrassegnate come riservate.

Microsoft afferma di aver rilasciato un aggiornamento per risolvere il problema e di "non aver fornito a nessuno l'accesso a informazioni che non erano già autorizzate a vedere".

Tuttavia, alcuni esperti hanno avvertito che la velocità con cui le aziende competono per aggiungere nuove funzionalità di intelligenza artificiale rende inevitabili questi tipi di errori.

Copilot Chat può essere utilizzato all'interno di programmi Microsoft come Outlook e Teams, per le e-mail e le funzioni di chat, per ottenere risposte a domande o riassumere messaggi.

"Abbiamo identificato e risolto un problema per cui Microsoft 365 Copilot Chat poteva restituire contenuti da e-mail etichettate come riservate, scritte da un utente e archiviate nelle bozze e negli elementi inviati sul desktop di Outlook", ha dichiarato un portavoce di Microsoft a BBC News.

"Sebbene i nostri controlli di accesso e le nostre politiche di protezione dei dati siano rimasti intatti, questo comportamento non soddisfaceva l'esperienza Copilot da noi prevista, che è progettata per escludere i contenuti protetti dall'accesso a Copilot", hanno aggiunto.

"Un aggiornamento della configurazione è stato distribuito in tutto il mondo per i clienti aziendali."

L'errore è stato segnalato per la prima volta dal sito di notizie tecnologiche **Bleeping Computer** , che ha affermato di aver visto un avviso di servizio che confermava il problema.

Citava un avviso di Microsoft in cui si affermava che "i messaggi di posta elettronica degli utenti con un'etichetta riservata vengono elaborati in modo errato dalla chat di Microsoft 365 Copilot".

L'avviso aggiungeva che una scheda di lavoro all'interno di Copilot Chat aveva riepilogato i messaggi di posta elettronica archiviati nelle bozze e nelle cartelle inviate di un utente, anche quando erano dotati di un'etichetta di riservatezza e di una politica di prevenzione della perdita di dati configurata per impedire la condivisione non autorizzata dei dati.

Secondo alcune indiscrezioni, Microsoft sarebbe venuta a conoscenza dell'errore per la prima volta a gennaio.

La segnalazione del bug è stata condivisa anche su una dashboard di supporto per gli operatori sanitari del Servizio Sanitario Nazionale in Inghilterra, dove la causa principale è attribuita a un "problema di codice".

Una sezione **dell'avviso sul sito di supporto IT dell'NHS** lascia intendere che è stato interessato.

Tuttavia, ha dichiarato alla BBC News che il contenuto di qualsiasi bozza o e-mail inviata ed elaborata da Copilot Chat resterà ai rispettivi creatori e che le informazioni sui pazienti non saranno divulgate.

"La fuga di dati avverrà"

Gli strumenti di intelligenza artificiale aziendali come Microsoft 365 Copilot Chat, disponibili per le organizzazioni con un abbonamento a Microsoft 365, spesso prevedono controlli e protezioni di sicurezza più rigorosi per impedire la condivisione di informazioni aziendali sensibili.

Tuttavia, per alcuni esperti, la questione evidenzia ancora i rischi dell'adozione di strumenti di intelligenza artificiale generativa in determinati ambienti di lavoro.

Nader Henein, analista di Gartner specializzato in protezione dei dati e governance dell'IA, ha affermato che "questo tipo di errore è inevitabile", data la frequenza con cui vengono rilasciate "nuove e innovative funzionalità di IA".

Ha detto alla BBC News che le organizzazioni che utilizzano questi prodotti di intelligenza artificiale spesso non dispongono degli strumenti necessari per proteggersi e gestire ogni nuova funzionalità.

"In circostanze normali, le organizzazioni semplicemente disattiverebbero la funzionalità e aspetterebbero che la governance si adeguasse", ha affermato Henein.

"Purtroppo la pressione esercitata dal flusso di propaganda infondata sull'intelligenza artificiale rende tutto ciò quasi impossibile", ha aggiunto.

L'esperto di sicurezza informatica, il professor Alan Woodward dell'Università del Surrey, ha affermato che ciò dimostra l'importanza di rendere tali strumenti privati per impostazione predefinita e accessibili solo tramite consenso esplicito.

"Inevitabilmente ci saranno dei bug in questi strumenti, soprattutto perché stanno avanzando a una velocità vertiginosa, quindi anche se la fuga di dati potrebbe non essere intenzionale, accadrà", ha detto a BBC News.