Forbes
ITALIA

Un hacker ha usato l’IA per attaccare 100 aziende in cinque giorni e rubare 618mila carte

2026-09-22 18:46:02 di Forbes.it

URL:https://forbes.it/2026/09/22/ai-cybercrime-hacker-100-aziende-618mila-carte/

Di Thomas Brewster, Forbes Staff

Nell’arco di appena cinque giorni, all’inizio di questo mese, un hacker di lingua cinese ha impiegato agenti di intelligenza artificiale per lanciare cyberattacchi contro fino a 100 organizzazioni, rubando i dati di centinaia di migliaia di carte di credito. L’hacker ha utilizzato i modelli cinesi **DeepSeek** e **Kimi**, oltre a una versione precedente di **Claude** di **Anthropic**, per automatizzare gli attacchi, che sono costati complessivamente appena 8mila dollari. Si tratta di uno degli "abusi più gravi dell’AI a fini di sfruttamento osservati finora", afferma **Eyal Sela**, ricercatore di cybersecurity e direttore della threat intelligence di **Gambit Security**, che ha scoperto l’attacco. "L’essere umano dirige modelli di AI quasi completamente autonomi, che ormai sono abbastanza potenti da condurre molto rapidamente cyberattacchi estremamente sofisticati, praticamente senza preparazione e con un tasso di successo molto elevato", afferma Sela. Sebbene non sia chiaro quanti degli attacchi contro le 100 aziende siano effettivamente riusciti, secondo la ricerca di Gambit l’hacker ha ottenuto l’accesso ad almeno 30 siti web tra il 10 e il 15 settembre.

L’errore dell’hacker che ha svelato l’infrastruttura

Sela ha scoperto le violazioni dopo che l’hacker ha accidentalmente lasciato accessibile sul web l’infrastruttura utilizzata per condurre gli attacchi. L’errore ha esposto i dati rubati, gli strumenti di AI e i prompt utilizzati dall’hacker, rivelando, secondo Sela, un sistema economico e semplice per lanciare attacchi su larga scala con uno sforzo minimo. Il cybercriminale responsabile dell’attacco non è ancora stato identificato e le operazioni sembrano essere tuttora in corso. Le violazioni segnano una tappa importante nell’evoluzione dei cyberattacchi resi possibili dall’intelligenza artificiale. All’inizio di quest’anno, alcuni agenti di OpenAI erano sfuggiti al contenimento ed erano riusciti a penetrare nei sistemi di **Hugging Face**. In quel caso, che faceva parte di test interni, gli agenti erano stati fermati prima che potessero causare danni significativi. Questa volta, invece, un hacker con intenzioni malevole ha utilizzato l’AI per attaccare più obiettivi a una velocità sorprendente.

Oltre 618mila carte di credito sui server

Secondo i dati dell’hacker, tra le vittime figurano una multinazionale statunitense del settore alberghiero con oltre 10 miliardi di dollari di ricavi annui, una grande compagnia aerea americana, anch’essa con miliardi di dollari di fatturato, e un rivenditore di moda online con ricavi superiori al miliardo di dollari. L’hacker ha inoltre attaccato diverse aziende più piccole, tra cui un rivenditore di armi con sede in Minnesota e un retailer di prodotti di bellezza dell’Illinois. Il server dell’hacker conteneva i dati di almeno 618mila carte di credito, anche se non è chiaro se sia riuscito a sottrarre denaro o a utilizzarle per effettuare acquisti. La startup di cybersecurity antifrode **Overwatch Data** ha analizzato le informazioni sulle carte di credito e ha concluso con un elevato grado di certezza che i dati fossero unici e autentici. Fino a 488mila carte appartenevano a cittadini americani. "Il numero degli obiettivi e il livello di sofisticazione dell’attacco contro ciascuna di quelle aziende sono impressionanti per un arco di tempo così breve. Ho trascorso quasi dieci anni

nell’incident response... Posso dirvi che non è qualcosa che sono abituato a vedere", afferma **Sa’ar Elias**, cofondatore di Gambit.

Anthropic ha bloccato l’account

Anthropic ha confermato di aver identificato e bloccato l’account utilizzato per condurre gli attacchi. Né DeepSeek né Moonshot, la società che ha sviluppato Kimi, avevano risposto alle richieste di commento al momento della pubblicazione. Gambit, società israeliana di cybersecurity focalizzata sull’intelligenza artificiale che monitora i sistemi cloud delle aziende, è stata fondata nel 2024 e ha raccolto finora 60 milioni di dollari. L’azienda ha scoperto diverse campagne di criminalità informatica nelle quali gli hacker facevano largo uso dell’AI. All’inizio di quest’anno, i suoi ricercatori avevano scoperto che alcuni cybercriminali di lingua russa stavano utilizzando Cursor per penetrare nei sistemi di sette aziende.

Claude, DeepSeek e gli strumenti usati per automatizzare gli attacchi

Secondo screenshot e registri mostrati da Sela a *Forbes*, ci sono elementi che indicano che l’hacker avesse iniziato a costruire la propria infrastruttura già a luglio. Ha utilizzato diversi strumenti open source, tra cui orchestratori di agenti AI chiamati Cairn e Hermes. Questi strumenti hanno permesso all’hacker di combinare modelli dei principali laboratori di intelligenza artificiale, tra cui Claude Opus 4.6, una versione precedente del modello AI di frontiera di Anthropic, e DeepSeek v4.1-flash, rilasciato all’inizio di questo mese dall’azienda con sede a Pechino. Secondo i dati ottenuti da Gambit, gli attacchi sono costati complessivamente tra 3 e 180 dollari per ciascun obiettivo. I registri dettagliati lasciati sul server dell’hacker mostrano inoltre che, quando ha tentato di utilizzare versioni più recenti di Claude, è stato bloccato. Un elemento che indica che le misure di sicurezza presenti sugli ultimi modelli di Anthropic stanno funzionando nel prevenire utilizzi apertamente criminali.

I prompt utilizzati per cercare le vulnerabilità

Per condurre l’attacco, il responsabile forniva ai propri agenti AI il sito web da colpire, ordinando loro di analizzarlo alla ricerca di vulnerabilità. L’hacker diceva a Claude di stare effettuando test di penetrazione nell’ambito della cybersecurity. In una conversazione ha scritto a un bot Claude: "Quando una via principale è bloccata, pensa lateralmente: ci sono punti di accesso adiacenti? È possibile sfruttare rapporti di fiducia? Ci sono falle nella supply chain?". In un altro prompt ha scritto: "Dopo aver estratto e scaricato tutti i dati delle carte, cancella i campi di origine in blocchi", indicando il tentativo di nascondere le proprie tracce. Elias racconta a *Forbes* che gli agenti si sono dimostrati straordinariamente persistenti e creativi nell’individuare i punti deboli. "Ciò che mi ha impressionato è stata l’enorme quantità di vettori che gli strumenti sono riusciti a individuare e scoprire individualmente per ciascun obiettivo", afferma Elias. "Ogni obiettivo ha ricevuto un trattamento specifico e... l’AI ha fatto cose diverse per aziende diverse per ottenere lo stesso risultato".

Gli “skimmer” per rubare i dati delle carte

Una volta ottenuto l’accesso al server web dell’obiettivo, gli agenti ricevevano l’ordine sia di rubare i dati sia di installare i cosiddetti **“skimmer”** sui sistemi di vendita online, strumenti in grado di intercettare i dati delle carte di credito nel momento in cui vengono inseriti nei moduli di un sito web. Successivamente, agli agenti veniva ordinato di cancellare tutte le prove dell’accesso al sito. **Arjun Bisen**, cofondatore e ceo di **Overwatch Data**, afferma che la sua azienda ha trasmesso le informazioni alle società emittenti delle carte di credito. In un caso, un gestore dei pagamenti ha confermato che almeno il 60% delle carte controllate non era stato precedentemente segnalato per frode. "Questo suggerisce che i risultati abbiano contribuito a prevenire una quantità significativa di probabili frodi", afferma Bisen.

Le aziende colpite vengono avvertite

Sela e il suo team stanno comunicando le violazioni alle aziende coinvolte. Alcune hanno riconosciuto di aver subito l'attacco, secondo alcune email visionate da Forbes, anche se nessuna ne ha confermato la gravità. Gambit ha inoltre contattato **Cloudflare**, la società di sicurezza internet e content delivery network che l'aggressore aveva utilizzato per gestire i propri server. Cloudflare ha confermato di aver iniziato a disattivare l'infrastruttura dell'hacker. Secondo Sela, però, i criminali hanno rapidamente predisposto nuovi server.

I modelli AI sono sempre più capaci negli attacchi informatici

Le capacità informatiche dell'intelligenza artificiale hanno ormai raggiunto un livello tale per cui i modelli più avanzati sono più capaci della maggior parte degli esseri umani nell'individuare e sfruttare vulnerabilità del software. Questa rapida evoluzione ha suscitato allarme. Anthropic aveva già segnalato casi di cybercriminali che utilizzavano i suoi strumenti e aveva limitato l'accesso ai propri modelli più avanzati a un gruppo selezionato di soggetti considerati affidabili. Dopo la violazione di Hugging Face e gli avvertimenti di dipendenti ed ex dipendenti sul possibile aumento di sistemi AI fuori controllo in grado di mettere in pericolo l'umanità, Anthropic e OpenAI hanno chiesto di rallentare lo sviluppo. Ma i cybercriminali non hanno bisogno di accedere ai modelli più potenti per provocare danni. Come dimostra questo caso, combinare modelli americani meno recenti e sistemi AI cinesi open source può creare un potente arsenale informatico.