La partita tra chi attacca e chi difende decisa tra codici e scrittura dei software

Le minacce. Solo OpenAi ha bloccato nel 2024 più di venti operazioni nel mondo che hanno utilizzato l'intelligenza artificiale generativa per campagne di disinformazione o tentare attacchi mirati sempre più precisi con una grammatica impeccabile

Enrico Pagliarini

In un rapporto datato ottobre 2024 OpenAI dichiara di «aver bloccato da inizio anno più di venti operazioni in tutto il mondo» che hanno usato, o tentato di usare, i modelli di intelligenza artificiale generativa del creatore di ChatGPT per produrre grandi quantità di contenuti per campagne di disinformazione oppure per attacchi mirati (o spear phishing) sempre più precisi, personalizzati e scritti con una grammatica impeccabile (in un buon italiano o in qualsiasi altra lingua). E questo è solo quello che la squadra di OpenAI è riuscita ad intercettare in una attività, quella di identificazione e analisi dei comportamenti sospetti che, si legge sempre nel rapporto, «sebbene richieda ancora un'intensa attività di valutazione e competenza umana si avvale di strumenti di intelligenza artificiale che hanno permesso di ridurre alcune fasi analitiche da giorni a minuti».

L'intelligenza artificiale, soprattutto quella generativa, è dunque uno straordinario strumento che la tecnologia ha dato ai criminali, ma sta fornendo benefici anche a chi deve difenderci. Chi ne stia beneficiando di più è oggetto di dibattito: ad inizio anno il 56% di un gruppo di leader interrogati dal World Economic Forum pensava che nei prossimi due anni «l'Ai generativa avvantaggerà i cyberattaccanti rispetto ai difensori». Vedremo.

Di certo il fenomeno è in crescita e destinato ad espandersi perché la diffusione della tecnologia ha reso oggi disponibili strumenti open source che le organizzazioni criminali possono usare per creare in casa i propri modelli linguistici di grandi dimensioni (o Llm) affrancandosi da servizi online come OpenAI e sfuggendo così ai controlli che bloccano sul nascere attività illecite. Gli scenari di utilizzo dell'intelligenza artificiale generativa sono diversi: campagne di disinformazione più ampie, mail di phishing più credibili e personalizzate che aprono la strada al ransomware, messaggi o chat che sembrano provenire da una persona di fiducia o da un'azienda legittima, truffe più efficaci con uso di voci sintetiche o cloni digitali (deepfake) che simulano una interazione verosimile con una persona fino a scenari in cui «chi attacca cerca di compromettere gli stessi sistemi di intelligenza artificiale», spiega Corrado Giustozzi, esperto di cybersecurity e partner di Rexilience, il quale invita a non sottovalutare la possibilità che «sistemi basati su Ia possano essere vulnerabili e quindi ingannati».

> **L'Ai sta alzando il livello dello scontro, aiutando sia i cybercriminali sia chi deve proteggere persone e aziende**

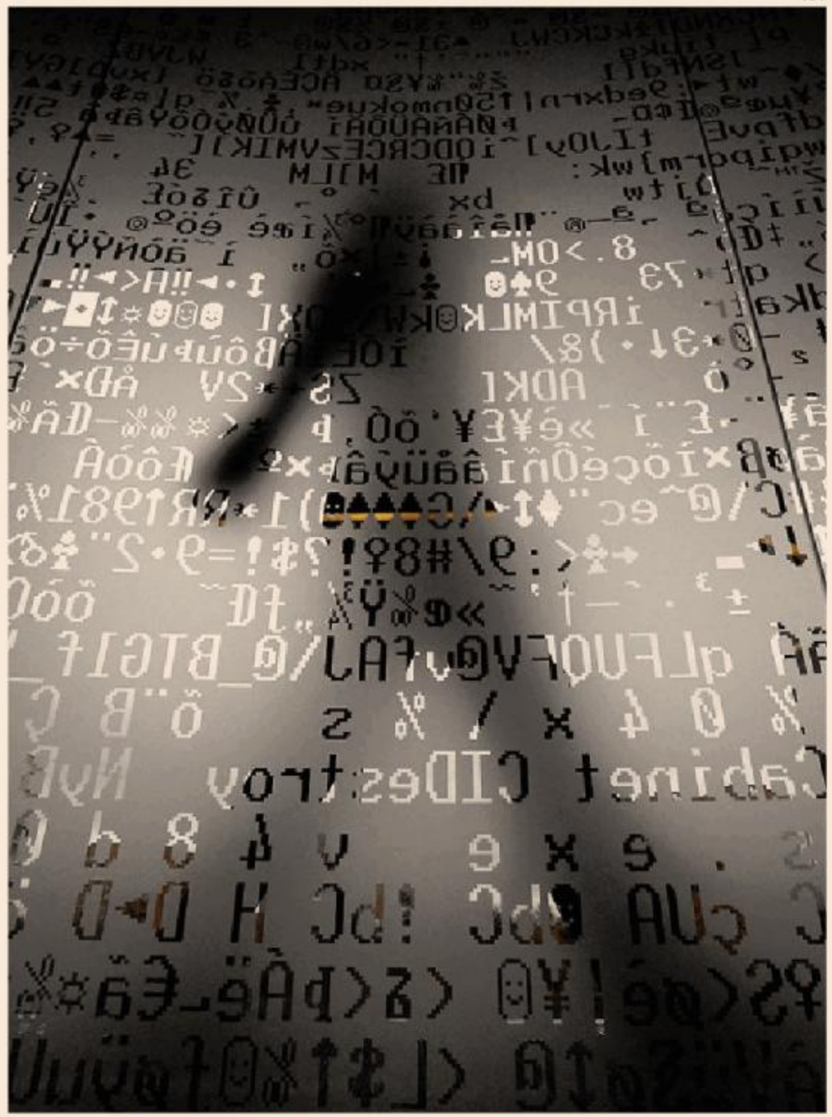
Nonostante il rapido cambiamento a cui stiamo assistendo, accelerato negli ultimi due anni proprio dall'Ia generativa «è opportuno sottolineare – prosegue Giustozzi – che di intelligenza artificiale e cybersecurity se ne parla da almeno 15 anni o più e in effetti abbiamo assistito alla nascita e diffusione di sistemi che cercano di cogliere segnali dalla rete per intercettare anomalie e reagire in modo automatico, sempre più velocemente e su enormi quantità di dati» non umanamente gestibili anche se «il fattore umano rimane fondamentale perché nessuno di questi sistemi può prendere decisioni in modo completamente autonomo».

La componente generativa può essere d'aiuto ai team di sicurezza che devono fare scelte operative in poco tempo perché «riassume oggi in un testo di poche righe le informazioni utili per prendere decisioni», aggiunge David Gubiani, regional director Se Emea Southern di Check Point Software Technologies. Le informazioni raccolte dalle piattaforme di cybersecurity rimangono ovviamente a disposizione, ma la macchina può generare suggerimenti in linguaggio naturale.

Un altro strumento usato da entrambe le parti, dai criminali che attaccano e da chi sviluppa software per difendere i sistemi digitali, è la capacità dei modelli di Ia generativa di scrivere software e di conseguenza anche malware. Con la GenAi è possibile generare codice nuovo e unico in quantità industriali e rendere il malware più difficile da identificare testando varianti del codice al fine di trovare la versione più efficace e invisibile ai sistemi di difesa. Ma a loro volta, in una rincorsa infinita che ripete uno schema già noto nel mondo della cybersecurity, i sistemi di difesa possono essere addestrati per prevenire in modo più efficace le minacce. «L'automazione nella scrittura del software è un vantaggio – aggiunge Gubiani – perché può ridurre il problema delle competenze e la mancanza di professionisti» che rimangono fondamentali ma rimarranno rilevanti nel mondo del lavoro nella misura in cui si adatteranno al nuovo scenario. «Sarà meno importante scrivere software ma – dice Gubiani – dovrai diventare bravo a dare ordini precisi ai sistemi di Ia per scrivere il codice».

C'è infine un aspetto che rimane immutato. La scarsa competenza informatica di base di molti lavoratori soprattutto nelle Pmi. «Purtroppo, ancora oggi i criminali informatici hanno un vantaggio culturale e a loro interessa che ci sia una buona parte di aziende compromissibili», conclude David Gubiani. Ed è anche in campo formativo che l'Ia generativa può dare un contributo con sistemi in grado di coinvolgere maggiormente i lavoratori. Al Sole 24 Ore, ad esempio, sono in corso sperimentazioni su percorsi formativi sulla cybersecurity erogati da un chatbot.

AFP



La tendenza. La diffusione dell'Ai facilita la proliferazione degli attacchi informatici