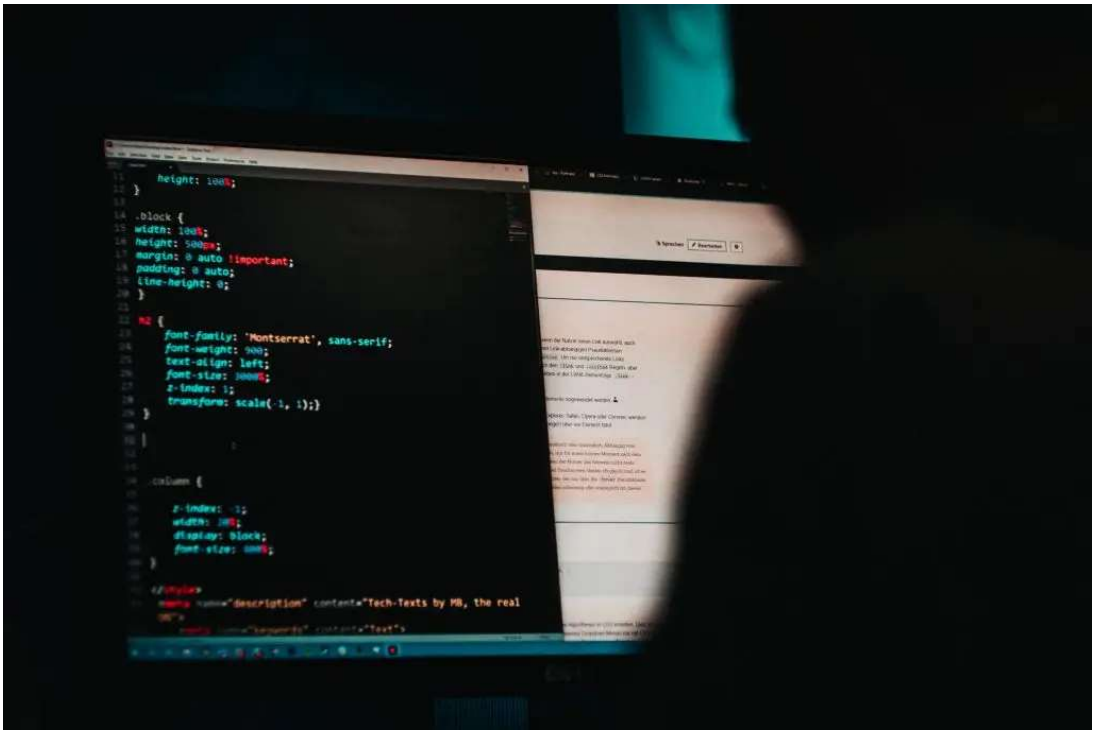
Chatbot ChatGPT: il lato oscuro dell'Intelligenza Artificiale

Da **Redazione BitMAT** - 19/01/2023

Ermes – Cybersecurity spiega quali sono i rischi di sicurezza digitale che il chatbot ChatGPT potrebbe comportare.



È sulla bocca di tutti, dopo il tanto atteso lancio a Novembre 2022: il chatbot **ChatGPT, acronimo di Chat Generative Pre-trained Transformer, è un chatbot basato sull'intelligenza artificiale, utilizzabile al momento gratuitamente** e creato dalla società **American OpenAI**, l'organizzazione non-profit di ricerca sull'intelligenza artificiale che promuove lo sviluppo delle AI amichevoli, ossia intelligenze capaci di contribuire al bene dell'umanità. Accedendo infatti al loro sito web, è possibile **conversare virtualmente con una "persona virtuale", un'intelligenza artificiale programmata per rispondere a qualsiasi quesito**, grazie ad un sofisticato modello di machine learning che possiede un'alta capacità di apprendimento automatico.

Quali sono, però, i rischi che il chatbot ChatGPT può comportare?

ChatGPT ha già attirato a sé molti cyber criminali, che in prima battuta hanno realizzato delle copie pressoché identiche del sito o dell'app, scaricabili dagli store ufficiali, che installate nel cellulare del malcapitato diffondono contenuti malevoli. Il problema più grave è però un altro: attraverso query specifiche e costruite ad arte, ChatGPT è il perfetto strumento che, nelle mani di un malintenzionato, lo aiuta a realizzare quelle che, nel mondo cyber, vengono chiamati attacchi di spear phishing. Si tratta, infatti, di attacchi iper customizzati, calibrati sulle informazioni che gli utenti, [illegible]a accorgersene, condividono sui loro account social e attraverso la navigazione

quotidiana su pc e mobile. In questo modo i cyber criminali utilizzano quindi l'AI per costruire un contenuto ingannevole, creato ad hoc per la persona a cui si rivolgono.

Ermes – Cybersecurity, specialista di Cybersicurezza, ha individuato i tre principali fattori di rischio legati al chatbot ChatGPT:

1. La truffa numero uno, quindi, è **la nascita di siti di phishing che sfruttano l'hype su ChatGPT,** già centinaia solo nelle ultime settimane. Riconoscerli non è facilissimo: hanno domini simili, aspetto delle pagine web o app pressoché identico e fanno spesso leva su integrazioni inesistenti, creando dei duplicati del servizio che rubano, così, le credenziali a tutti coloro che vi si registrano;
2. **Gli attacchi di spear phishing** diventano sempre più facili e scalabili con la produzione qualitativa e veloce di campagne email (BEC), sms (smishing) o ads (malaware) estremamente targetizzate, volte a truffe economiche, furto di dati personali o credenziali;
3. **La condivisione di informazioni sensibili dell'azienda**, con la richiesta continua di contenuti, risposte ed analisi. Come avviene questo? Ad esempio con un semplice "rispondi a questa mail" dimenticandosi di escludere l'email del destinatario o del mittente, oppure dando in pasto a queste nuove tecnologie dati economici o nomi di clienti o partner.

Un esempio pratico: Business Email Compromise, il rischio per le mail aziendali.

ChatGPT risponde in maniera eccellente a qualsiasi query di contenuti, ma **ciò diventa particolarmente rischioso se utilizzato come attacco delle e-mail aziendali**, il cosiddetto BEC. Con BEC, gli aggressori utilizzano un modello per generare un'email ingannevole, che spinge un destinatario a fornire lui informazioni sensibili. Con l'aiuto di ChatGPT, infatti, gli hacker avrebbero la possibilità di customizzare ogni comunicazione, avendo così potenzialmente contenuti unici per ogni e-mail generata grazie all'AI, rendendo questi attacchi più difficili da rilevare e riconoscere come tali.

Allo stesso modo, scrivere e-mail o costruire il copy di un sito di phishing può diventare più facile, senza errori di battitura o formati unici, che oggi sono spesso critici per differenziare questi attacchi da e-mail legittime. Ciò che più spaventa è che diventa così possibile aggiungere il maggior numero di variazioni al prompt, come "rendere l'email urgente", "e-mail con un'alta probabilità di destinatari che fanno clic sul link" e così via.

Redazione BitMAT

https://www.bitmat.it/

BitMAT Edizioni è una casa editrice che ha sede a Milano con una copertura a 360° per quanto riguarda la comunicazione rivolta agli specialisti dell'Information & Communication Technology.