ChatGPT e cybercrime: l’analisi dell’Europol



Già dopo pochi mesi dal suo rilascio sul mercato, avvenuto nel novembre 2022, l’impatto di ChatGPT risulta dirompente.

Il Large Language Model (LLM) sviluppato da OpenAI permette di velocizzare e potenziare moltissimi flussi di lavoro “leciti”, tra cui attività di ricerca o traduzione di contenuti: ma ad allarmare è il suo possibile utilizzo a fini criminali, specie considerando come possa ridurre le barriere d’ingresso al mercato del *cybercrime* per chi intenda sferrare attacchi basati sull’ingegneria sociale pur senza avere particolari competenze informatiche.

Per questo l’**Europol Innovation Lab** – che annovera tra i propri compiti il monitoraggio delle intersezioni tra tecnologie ed ecosistemi criminali, nonché la proposta di soluzioni innovative in materia agli Stati UE – negli ultimi mesi ha svolto una serie di *workshop* per esaminare i potenziali abusi del nuovo strumento.

La ricerca ha contemplato anche il “lato positivo”, coinvolgendo esperti di vari ambiti al fine di approfondire le potenzialità di ChatGPT come risorsa per indagini digitali, operazioni di *counterterrorism* o attività rientranti nel più generale campo dell’*information technology*; così confermando la natura di “arma a doppio taglio” dei LLMs e nello specifico di ChatGPT, scelta in qualità di oggetto dell’analisi per via della sua forte diffusione nel panorama attuale.

Il risultato è il report di sintesi *“Tech Watch Flash: The impact of Large Language Models on Law Enforcement”*, accompagnato da una più ampia versione ad uso esclusivo di forze dell’ordine e organismi investigativi.

Partendo dalle definizioni di *Artificial Intelligence* (AI), *Artificial Neural Networks* (ANN) e *Deep Learning* fornite dalla stessa ChatGPT, il testo ne ripercorre genesi e fasi di sviluppo per poi valutare i limiti mostrati dalla tecnologia in esame; fra di essi la scarsa trasparenza circa il trattamento riservato ai dati personali, il fatto che buona parte dei *dataset* usati per "allenarla" risalgano al settembre 2021 e la tendenza a fornire risposte plausibili, ma non sempre corrette, alle domande degli utenti.

Molto concrete appaiono le preoccupazioni relative alle attività di *social engineering*, legate soprattutto al fatto che ChatGPT potrebbe facilitare l'impersonificazione di persone ignare abilitando, così, campagne di *phishing* sofisticate e su larga scala.

Si evidenzia poi il rischio che i meccanismi di sicurezza previsti da OpenAI vengano aggirati da utenti malevoli tramite il *prompt engineering*: in particolare grazie al comando DAN (*"Do Anything Now"*), specificamente pensato per bypassare i criteri di moderazione inducendo il sistema a rispondere a qualsiasi input successivo.

Il report cita anche possibili utilizzi a fini di disinformazione e propaganda, oltre naturalmente all'ipotesi (già verificatasi) che un *threat actor* si serva di Chat GPT per ottenere codice malevolo con cui lanciare attacchi informatici mirati agli obiettivi più svariati, ovvero per commettere reati gravissimi quale lo sfruttamento sessuale di minori.

Sebbene il recente rilascio di ChatGPT-4 prometta di superare questi limiti, rimangono diversi punti oscuri che andranno attentamente monitorati per evitare che – come già accaduto in passato – l'innovazione tecnologica si tramuti in un vantaggio per la criminalità informatica.

Il report dell'Europol termina, infatti, con una serie di

raccomandazioni che intimano alla *law enforcement community* di non abbassare la guardia sui rischi evidenziati; invitando altresì a promuovere formazione e cooperazione interforze per sfruttare i potenziali benefici di ChatGPT, senza rinunciare a prevenire ogni possibile abuso.

Condividi sui Social Network:
https://www.ictsecuritymagazine.com/notizie/chatgpt-e-cybercrime-lanalisi-delleuropol/