L'Europol "indaga" su ChatGPT. L'IA facilita le nuove frontiere del crimine

di Alessia Palladino - pubblicato su "www.irpa.eu" - Osservatorio sullo Stato digitale, 31 maggio 2023

Lo sviluppo di ChatGPT può rappresentare un rischio per il rispetto dei diritti fondamentali e per lo sviluppo della criminalità. Per tali ragioni, l'Europol Innovation Lab ha organizzato una serie di workshop per valutare in che modo i criminali possono abusare di ChatGPT.

Lo sviluppo dell'**Intelligenza artificiale** e dei **Large Language Model (LLM)** hanno guadagnato nuova popolarità attraverso la diffusione di **ChatGPT**, un modello di linguaggio di grandi dimensioni (LLM) sviluppato da OpenAI già nel 2020.

Il successo di ChatGPT risiede nell'utilizzo di un **algoritmo di Deep learning**, capace di analizzare i set di Big data forniti per riconoscere contenuti, generarli, riassumerli, tradurli nonché prevederli. ChatGPT assimila le capacità di un modello linguistico a quelle di un essere umano capace di fornire risposte immediate ed estremamente versatili, tali da poter essere applicate a una vasta quantità di contesti diversi (per approfondimenti si rimanda a *B. Carotti, Nulla di nuovo sul fronte artificiale: ChatGPT e gli altri*).

In questo senso, ChatGPT **pone le basi per l'impiego creativo dell'Intelligenza Artificiale** e dell'apprendimento automatico, non più esclusivamente utilizzato per gestire compiti banali, a carattere seriale, dimostrandosi capace di lavori creativi complessi.

Sebbene lo sviluppo di ChatGPT e dei LLM offra grandi opportunità di natura economico – sociale, al contempo esso può rappresentare un rischio per il rispetto dei diritti fondamentali e per lo sviluppo della criminalità.

Per tali ragioni, **l'Europol Innovation Lab** ha organizzato una serie di workshop con esperti in materia, che rappresentavano l'intero impianto delle competenze di Europol (compresa l'analisi operativa, la criminalità organizzata, la criminalità informatica, l'antiterrorismo e la tecnologia dell'informazione), per valutare in che modo i criminali potrebbero abusare di tecnologie come ChatGPT, nonché come tale tecnologia possa costituire un supporto per le attività investigative.

Il rapporto **"*Tech Watch Flash*"** – ***ChatGPT – The impact of Large Language Models on Law Enforcement, a Tech Watch Flash Report from the Europol Innovation Lab,*** analizza i risultati di queste sessioni e include informazioni chiave per le forze dell'ordine.

In particolare, i workshop di Europol hanno identificato una **vasta gamma di casi di uso criminale** in GPT: com'è noto, infatti, ChatGPT eccelle nel fornire all'utente informazioni in risposta a un'ampia gamma di richieste.

Pertanto, **se un potenziale criminale non dispone di conoscenze per una particolare area criminale, ChatGPT può accelerare notevolmente il processo di ricerca** offrendo informazioni chiave che possono essere successivamente approfondite. Più in dettaglio, ChatGPT può essere utilizzato per acquisire conoscenza sugli atti preparatori, idonei a

commettere un reato, illustrando le migliori modalità per entrare in una casa, nonché facilitando le attività di terrorismo, criminalità informatica e abusi sessuali su minori.

I casi d'uso emersi dai workshop svolti da Europol non sono affatto esaustivi: piuttosto, l'obiettivo è di fornire un'idea di quanto variegati e potenzialmente pericolosi possano essere tali strumenti nelle mani di malintenzionati.

In particolare, i principali settori di analisi, confluiti nel Report sono essenzialmente due:

- **Frode, impersonificazione e *social engineering***;
- **Cybercrimes**.

Frode, impersonificazione e social engineering. ChatGPT può offrire ai criminali nuove opportunità, specialmente per quei crimini che coinvolgono il *social engineering*, grazie alla sua capacità di rispondere ai messaggi nel contesto e di adottare uno stile di scrittura specifico.

In particolare, **la capacità di ChatGPT di redigere testi altamente autentici**, sulla base di un *prompt* dell'utente, **lo rende uno strumento estremamente utile per le attività di *phishing***. Peraltro, se i tentativi di truffa attraverso il phishing potevano essere in precedenza più facilmente rilevabili, a causa di evidenti errori grammaticali e ortografici, le capacità linguistiche di ChatGPT consentono di abbattere tali criticità, impersonando un'organizzazione o un individuo in modo altamente realistico anche con una conoscenza di base della lingua inglese.

Con l'attuale versione di ChatGPT è già possibile creare strumenti di base per una varietà di scopi dannosi, ad esempio per produrre pagine di phishing o script VBA (Visual Basic for Applications) dannosi.

Cybercrimes. Una delle aree di criminalità su cui ciò potrebbe avere un impatto significativo è quello della criminalità informatica, perché consente a chiunque, sprovvisto di conoscenze tecniche, di sfruttare un vettore di attacco sul sistema di una vittima.

La capacità di ChatGPT di trasformare i *prompt* di linguaggio naturale in codice funzionante è stata rapidamente sfruttata da malintenzionati **per creare malware**.

In conclusione, il Report dell'Europol manifesta l'idoneità di ChatGPT a fungere da facilitatore per la commissione di crimini, stimolando l'attitudine a delinquere.

In risposta alle crescenti esigenze di garantire che i modelli di IA generativa siano sicuri, **Partnership on AI** (PAI), un'organizzazione di ricerca senza scopo di lucro, ha stabilito una serie di linee guida su come produrre e condividere i contenuti generati dall'IA in modo responsabile. Queste linee guida sono state sottoscritte da un gruppo di dieci aziende, tra cui OpenAI, impegnandosi ad aderire a una serie di best practice.

A livello europeo, invece, le istituzioni europee stanno finalizzando gli sforzi legislativi volti a regolamentare i sistemi di intelligenza artificiale. Sebbene il fenomeno non trovi esplicita regolamentazione, i sistemi di Intelligenza Artificiale come ChatGPT dovrebbero essere inclusi nell'ambito dei "sistemi ad alto rischio".