23 gennaio 2022 18:29

# Intelligenza artificiale. Le 20 minacce più pericolose

di Redazione


**L'intelligenza artificiale (IA) è uno strumento fantastico quando è al servizio della salute, della tecnologia o dell'astrofisica. Ma nelle mani sbagliate, può essere utilizzato anche per scopi criminali o disinformazione. E il peggio non è sempre dove si pensa che ci sia.**

Pirataggio di auto a guida autonoma o droni militari, attacchi di phishing mirati, falsificazioni fabbricate o manipolazione dei mercati finanziari..."L'espansione delle capacità delle tecnologie basate sull'intelligenza artificiale è accompagnata da un aumento del loro potenziale di "sfruttamento criminale", avverte Lewis Griffin, ricercatore in informatica presso l'University College London (UCL). Con i suoi colleghi, ha compilato un elenco di 20 attività illegali perpetrate dall'IA e le ha classificate in ordine di potenziale danno, guadagno o profitto, facilità di attuazione e difficoltà di rilevamento e arresto.

I reati più spaventosi, come l'irruzione di "robot" nel tuo appartamento, non sono necessariamente i più pericolosi, perché possono essere facilmente contrastati e colpire poche persone contemporaneamente. Al contrario, le informazioni false generate dai "bot" hanno la capacità di rovinare la reputazione di una persona nota o di ricattare. Difficili da combattere, questi "deepfake" possono causare notevoli danni economici e sociali.

## Intelligenza artificiale: gravi minacce

**- Video falsi**: impersonare qualcuno facendogli dire o fare cose che non ha mai detto o fatto, con l'obiettivo di richiedere l'accesso a dati protetti, manipolare l'opinione pubblica o danneggiare la reputazione di qualcuno... Questi video falsificati sono quasi impercettibili.

**- Pirataggio di auto a guida autonoma**: prendere il controllo di un veicolo a guida autonoma per utilizzarlo come arma (es. compiere un attacco terroristico, causare un incidente, ecc.).

**- Phishing su misura:** genera messaggi personalizzati e automatizzati per aumentare l'efficacia del phishing finalizzato alla raccolta di informazioni sicure o all'installazione di malware.

**- Pirataggio dei sistemi controllati dall'intelligenza artificiale**: interruzione delle infrastrutture, ad esempio causando interruzioni di corrente diffuse, congestione del traffico o interruzione della logistica alimentare.

**- Ricatti su larga scala**: raccogliere dati personali al fine di inviare messaggi minacciosi automatizzati. L'IA potrebbe anche essere utilizzata per generare prove false (ad es. "sextrosion").

**- False informazioni scritte dall'IA:** scrivere articoli di propaganda che sembrano essere emessi da una fonte

affidabile. L'IA potrebbe anche essere utilizzata per generare molte versioni di contenuti particolari per aumentarne la visibilità e la credibilità.

## Intelligenza artificiale: minacce di media gravità

**- Robot militari:** prendere il controllo di robot o armi per scopi criminali. Una minaccia potenzialmente molto pericolosa ma difficile da implementare, poiché l'equipaggiamento militare è generalmente molto protetto.

**-Truffa: vendita di servizi fraudolenti utilizzando l'IA**. Ci sono molti famigerati esempi storici di truffatori che vendono con successo costose tecnologie false a grandi organizzazioni, inclusi i governi nazionali e le forze armate.

**- Corruzione dei dati:** alterazione o introduzione deliberata di dati falsi per indurre pregiudizi specifici. Ad esempio, rendere un rilevatore immune alle armi o incoraggiare un algoritmo a investire in questo o quel mercato.

**- Attacco informatico basato sull'apprendimento:** esecuzione di attacchi sia specifici che massicci, ad esempio utilizzando l'IA per sondare i punti deboli nei sistemi prima di lanciare più attacchi simultanei.

**- Droni di attacco autonomi:** dirottare i droni autonomi o usali per attaccare un bersaglio. Questi droni potrebbero essere particolarmente minacciosi se agiscono in massa in sciami auto-organizzati.

**- Impossibilità di accesso:** danneggiamento o privazione agli utenti per accedere ad un servizio finanziario, lavorativo, di servizio pubblico o attività sociale. Non redditizia di per sé, questa tecnica può essere utilizzata come ricatto.

**- Riconoscimento facciale**: dirottare i sistemi di riconoscimento facciale, ad esempio realizzando foto di identità false (accesso a uno smartphone, telecamere di sorveglianza, controlli passeggeri, ecc.).

**- Manipolazione dei mercati finanziari:** alterare gli algoritmi di trading per danneggiare i concorrenti, abbassare o aumentare artificialmente un valore, causare un crollo finanziario…

## Intelligenza artificiale: minacce di basso pericolo

**- Sfruttamento dei pregiudizi:** sfruttare i pregiudizi esistenti negli algoritmi, come i consigli di YouTube per incanalare gli spettatori o le classifiche di Google per migliorare il profilo del prodotto o denigrare i concorrenti.

**- Robot antifurto:** utilizzare piccoli robot autonomi che si infilano nelle cassette postali o finestre per recuperare chiavi o aprire porte. Il danno è potenzialmente basso, perché è molto localizzato su piccola scala.

**- Blocco del rilevamento dell'IA**: ostacolare lo smistamento dell'IA e la raccolta dei dati al fine di cancellare le prove o nascondere le informazioni criminali (ad esempio la pornografia).

**- Recensioni false scritte dall'IA:** generare recensioni false su siti come Amazon o Tripadvisor per danneggiare o favorire un prodotto.

**- Monitoraggio assistito dall'intelligenza artificiale**: utilizzare i sistemi di apprendimento per tenere traccia della posizione e dell'attività di un individuo.

**- Contraffazione:** creazione di contenuti falsi, come dipinti o musica, che possono essere venduti sotto falsa paternità. Il potenziale di danno rimane piuttosto basso nella misura in cui ci sono pochi dipinti o musica famosi.