UNIVERSITÀ
POLITECNICA
DELLE MARCHE

DII
Dipartimento di Ingegneria
dell'Informazione

unimc

Dalla minaccia alla difesa: il ruolo dell'IA nella Cybersecurity

Paolo Costigliola

Venerdì 3 Ottobre 2025

Corso di Perfezionamento in Cybersecurity, Cyber Risk and Data Protection

Introduzione

La sicurezza informatica **evolve** costantemente, sostenuta dalla **nascita** e dalla **crescita** delle **tecnologie emergenti**.

Tra le tecnologie emergenti **più rilevanti** figura **l'intelligenza artificiale (IA),** che ha **trasformato** in modo **radicale** il **panorama** della **cybersecurity**.

Cos’ è la **Cybersecurity**?

Definizione

«Insieme delle misure **fisiche**, **logiche** e **procedurali**, finalizzate a **garantire riservatezza**, **integrità** e **disponibilità** delle **informazioni** elaborate tramite **sistemi informatici**.»

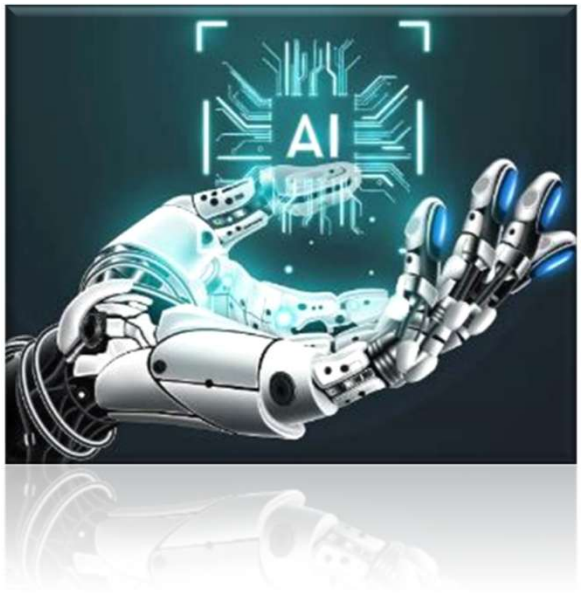
Fonte: Agenzia per la Cybersicurezza Nazionale (ACN)

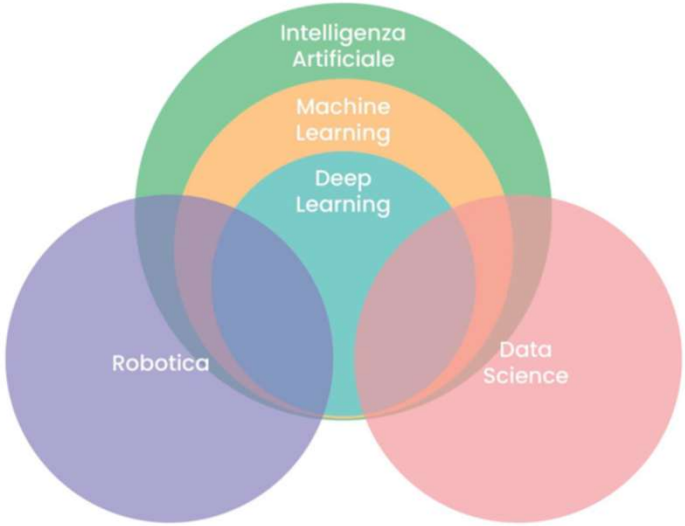
Cos'è l'**Intelligenza Artificiale**?

Definizione

«L'intelligenza artificiale (IA) è la **capacità** di un computer o di un robot controllato da esso di svolgere **compiti** comunemente associati alle **entità intelligenti**.»

Fonte: Encyclopedia Britannica –B.J. Copeland





Ruolo dell'**IA** nella **Cybersecurity**

Intelligenza Artificiale e Cybersecurity

Il panorama delle **minacce informatiche** evolve **costantemente**, con **tecniche sempre più sofisticate** che rendono la sicurezza informatica una **sfida sempre più complessa**.

In questo contesto, l'**IA** potrebbe configurarsi:

- uno **strumento/vettore di attacco**
- uno **strumento/vettore di difesa**
- un **bersaglio** o una «superficie» **da attaccare**

L'intelligenza artificiale come **STRUMENTO** d'**ATTACCO**

IA come strumento d’ attacco

L’applicazione di **tecniche e algoritmi avanzati** aumenta l’efficacia degli **attacchi informatici**.

Nasce così uno **squilibrio** tra chi attacca e chi difende. L’**IA** amplifica questo **squilibrio**, perché gli attaccanti diventano più **rapidi ed efficaci** nell’adattare e migliorare le loro tecniche.

Un ambito particolarmente colpito è quello dell’**ingegneria sociale**.

Social Engineering

L'ingegneria sociale consiste **nell'utilizzo di metodi** che mirano a ottenere informazioni da persone, aziende o enti tramite **inganno**.

Un aspetto strettamente correlato all'evoluzione delle tattiche di attacco è **l'automatizzazione degli attacchi di Social Engineering**, **potenziata dall'uso** dell'**intelligenza artificiale generativa**.

Permette di passare da attacchi massivi e generici a **attacchi altamente personalizzati**, aumentando il tasso di successo.

Altri utilizzi dell’intelligenza artificiale negli attacchi

- **Vishing:** Le chiamate non sono più robotiche e generiche: l’AI può generare voci sintetiche realistiche, imitare persone autorevoli o conosciute dalla vittima
- **Malware polimorfo:** modifica continuamente il codice per sfuggire agli antivirus.
- **Ransomware:** tecniche crittografiche avanzate e adattamento automatico rendono più difficile la rilevazione.
- **Attacchi biometrici:** volti o impronte digitali generati dall’AI possono ingannare sistemi di riconoscimento.
- **Agent AI (agenti autonomi):** sistemi intelligenti che individuano vulnerabilità, creano exploit e adattano il codice in autonomia, senza intervento umano, rendendo gli attacchi più rapidi ed efficaci.

L'intelligenza artificiale come **STRUMENTO** di **DIFESA**

Corso di Perfezionamento in Cybersecurity, Cyber Risk and Data Protection

Utilizzare l’IA per contrastare gli attacchi

Utilizzare **l’intelligenza artificiale** (IA) come **strumento di difesa** è ormai quasi **obbligatorio**, dato che gli attacchi informatici diventano sempre più **complessi e difficili** da **rilevare**.

Un ulteriore vantaggio è l’**automatizzazione** delle risposte, che consente di **isolare rapidamente** dispositivi infetti, **disattivare credenziali compromesse** e **applicare patch di sicurezza**.

L’IA a supporto dei 4 pilastri della sicurezza informatica

- **Sicurezza dei dati**: l’IA può rafforzare la protezione dei dati tramite configurazioni autonome, modificando i privilegi di accesso in caso di attacco e ripristinandoli quando la minaccia è cessata.
- **Sicurezza delle applicazioni**: l’IA supporta gli sviluppatori nella generazione di codice più sicuro e privo di bug, riducendo i rischi legati allo sviluppo software.
- **Sicurezza dell’identità**: l’IA consente metodi di autenticazione avanzati, come il riconoscimento facciale o le impronte digitali, diminuendo le vulnerabilità legate alle password.
- **Sicurezza di rete**: tecnologie come lo zero trust packet routing e l’analisi di miliardi di pacchetti quasi in tempo reale proteggono la rete da esposizioni e vulnerabilità.

L’IA nei Security Operations Center (SOC)

Un **Security Operations Center** (SOC) è un **centro di comando** centralizzato che **monitora, analizza** e **risponde** agli **incidenti** di **sicurezza informatica**.

EDR(Endpoint Detectionand Response)

NDR(Network Detectionand Response)

UEBA(User Entityand BehaviorAnalytics)

SIEM(Security Information and Event Management)

L'intelligenza artificiale come **SUPERFICIE** d'**ATTACCO**

L’IA come bersaglio

L’IA è sempre **più integrata in sistemi quotidiani** — come veicoli autonomi, riconoscimento facciale, filtri antispam e assistenti vocali — e, di conseguenza, offre una **superficie di attacco ampia**.

I **modelli** di **intelligenza artificiale** possono diventare un **obiettivo**: gli **attaccanti** possono **sfruttare** le loro **vulnerabilità** per causare **danni concreti**.

Tipologie di attacchi

I principali attacchi ai modelli di IA sono quattro:

- **Adversarial attack**: manipolazione degli input per ingannare il modello.
- **Data poisoning**: avvelenamento dei dati di addestramento per compromettere il modello.
- **Model stealing**: furto del modello o estrazione di informazioni sensibili.
- **Data leakage**: violazione della privacy dei dati utilizzati per l'addestramento.

Caso **GTG-2002**

GTG-2002

GTG-2002 (Claude Code) segna un **punto di svolta** nel panorama delle **minacce informatiche**.

IA **Claude Code** usata per la prima volta come **agente attivo** di **attacchi informatici**.

Dinamica dell'attacco.

Conclusioni

Sfide future

La **principale** è che queste tecnologie **continueranno a evolvere**, alimentando un **costante** braccio di ferro tra **attaccanti** e **difensori**.

L'IA **renderà possibile** automatizzare **molti tipi di attacchi**, inclusi quelli **innovativi o difficili da rilevare** dai sistemi difensivi tradizionali. Ciò comporta un **aumento** del livello **della minaccia**.

VI RINGRAZIO PER L'ATTENZIONE