UNIVERSITÀ POLITECNICA DELLE MARCHE

DII
Dipartimento di Ingegneria dell'Informazione

unimc

Intelligenza Artificiale e Cybersecurity: il triplice ruolo tra attacco, difesa e superficie di attacco

Matteo Alfieri

Dipartimento di Ingegneria dell'Informazione - DII

m.alfieri.ing@proton.me

www.linkedin.com/in/matteo-alfieri-ing

Giovedì 02 Ottobre 2025

Corso di Perfezionamento in Cybersecurity, Cyber Risk and Data Protection

Focus dell'elaborato

In un contesto in cui i sistemi informativi sono sempre più automatizzati, distribuiti e interconnessi, **l'intelligenza artificiale** rappresenta un **fattore trasformativo**, capace di potenziare le minacce informatiche e di rafforzare le capacità difensive.

Il presente elaborato intende:

- esaminare come l'IA venga impiegata in **tecniche offensive**, quali la generazione automatizzata di phishing, l'evasione dei sistemi di rilevamento e gli attacchi condotti da agenti autonomi;
- approfondire le **applicazioni difensive**, in particolare nell'ambito dei Security Operations Center (SOC), dell'analisi dei comportamenti anomali e dell'automazione delle risposte;
- analizzare l'IA come **superficie di attacco**, con attenzione a vettori emergenti quali il model stealing, gli attacchi adversariali e le implicazioni per la privacy dei dati.

L'obiettivo finale è offrire una panoramica tecnica dei **rischi e** delle **opportunità** che l'IA introduce nella sicurezza informatica, considerando anche gli aspetti giuridici ed etici. La multidisciplinarità del tema richiede infatti un **approccio olistico**.

Argomenti

- **1 Introduzione**
  - **Intelligenza Artificiale**
    - Definizione
    - Origine del termine
    - Ecosistema
  - **Cybersecurity**
    - Definizione
    - Triade CIA:
      (Riservatezza, Integrità e Disponibilità)
    - Importanza della Cybersecurity
- **2 Normative giuridiche di riferimento in ambito Cybersecurity e IA**
- **3 IA come strumento di attacco**
  - IA come forza moltiplicatrice per l'attacco
  - Evoluzione della Social Engineering con IA
  - Nuove frontiere dell'attacco:
    malware, agenti autonomi e contromisure
- **4 IA come strumento di difesa**
  - IA come strumento proattivo di Cyber Difesa
  - SOC e automazione delle risposte con IA
  - Difesa Passiva e Attiva basata su IA
- **5 IA come superficie di attacco**
  - Minacce e vettori
  - Furto di modelli e attacchi alla privacy
  - Casi studio, implicazioni e mitigazioni
- **6 L'etica nella IA e Cybersecurity**
- **7 Conclusioni**

Introduzione all’Intelligenza Artificiale

Corso di Perfezionamento in Cybersecurity, Cyber Risk and Data Protection

Introduzione all'Intelligenza Artificiale 1/2

- **Definizione:**

  L'Intelligenza Artificiale (IA) è la simulazione dei processi di intelligenza umana da parte di sistemi informatici.
  Permette a computer e macchine di simulare: apprendimento, comprensione, risoluzione di problemi, processo decisionale, creatività e una certa forma di autonomia operativa umana.

- **Origine del termine:**

  Il termine *"Intelligenza Artificiale*" fu coniato da **John McCarthy** nel **1955**, all'interno di una proposta di ricerca dedicata allo studio di macchine capaci di simulare l'intelligenza umana.

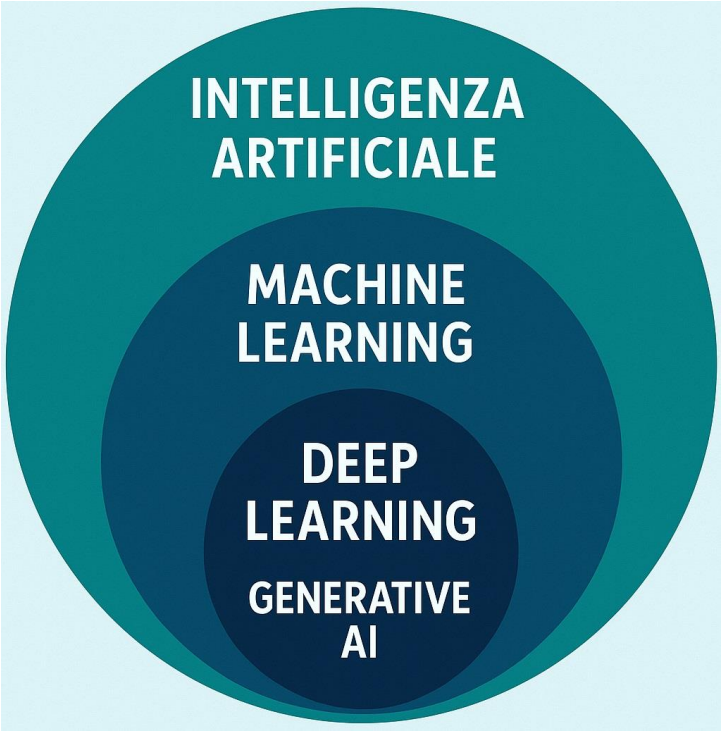
  L'anno successivo, nel 1956, il concetto fu presentato ufficialmente durante il Dartmouth Summer Research Project on Artificial Intelligence, tenutosi presso il Dartmouth College di Hanover (New Hampshire, USA), e promosso da John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon. L'evento è considerato l'atto di nascita della disciplina.

Introduzione all'Intelligenza Artificiale 2/2

- **Ecosistema dell'Intelligenza Artificiale:**

- **Machine Learning:**

  Il Machine Learning (ML) è un sottoinsieme dell'Intelligenza Artificiale che studia algoritmi in grado di migliorare attraverso l'esperienza. Può utilizzare reti neurali e si articola, principalmente, in apprendimento supervisionato e non supervisionato.



- **Deep Learning:**

  Il Deep Learning (DL) è un sottoinsieme del Machine Learning che studia algoritmi ispirati alla struttura e al funzionamento del cervello umano, chiamati reti neurali artificiali.

- **Generative AI:**

  È un ramo dell'Intelligenza Artificiale che si concentra sulla creazione di nuovi dati (testi, immagini, audio o video) a partire da dati esistenti. Utilizza modelli di Machine Learning per identificare schemi e relazioni nei dati di partenza, che vengono poi sfruttati per generare contenuti simili ma non identici agli originali. Esempio: Large Language Models (LLM).

Introduzione alla Cybersecurity

Corso di Perfezionamento in Cybersecurity, Cyber Risk and Data Protection

Introduzione alla Cybersecurity 1/2

- **Definizione di Cybersecurity:**
  È la disciplina che si occupa della protezione di sistemi informatici, reti e dati da attacchi malevoli, accessi non autorizzati e altre minacce digitali.
  Include misure tecniche, organizzative e procedurali volte a garantire la sicurezza delle informazioni in ambienti digitali sempre più complessi e interconnessi.

- **Triade CIA** (Confidentiality, Integrity e Availability),
  fondamento teorico della sicurezza delle informazioni.
  - **Riservatezza** (**C**onfidentiality):
    I dati devono essere accessibili solo da utenti, entità o processi autorizzati**.**
  - **Integrità** (**I**ntegrity):
    I dati devono mantenere la loro correttezza, completezza e affidabilità, evitando modifiche non autorizzate (volontarie o accidentali).
  - **Disponibilità** (**A**vailability):
    Le informazioni devono essere accessibili, nei tempi previsti, da parte degli utenti autorizzati.

Tutti e tre gli elementi devono essere bilanciati in modo efficace per garantire una sicurezza completa.

Introduzione alla Cybersecurity 2/2

- **Importanza della Cybersecurity:**

  Nel contesto attuale, caratterizzato da minacce sempre più evolute e automatizzate, la cybersecurity rappresenta un pilastro della resilienza digitale.

  Attacchi informatici possono generare danni significativi a individui, aziende e infrastrutture tra cui:

  - Perdita o furto di dati sensibili (personali, finanziari e aziendali);
  - Costi elevati di bonifica e ripristino;
  - Sanzioni economiche (ad es. dal Garante per la protezione dei dati personali in Italia);
  - Danni reputazionali;
  - Perdita di fiducia da parte di clienti e partner;
  - Interruzione delle attività aziendali (downtime e blocchi di produzione).

  La prevenzione non è un'opzione, ma una necessità strategica.

Normative giuridiche di riferimento in ambito Cybersecurity e IA

Corso di Perfezionamento in Cybersecurity, Cyber Risk and Data Protection

Normative giuridiche di riferimento in ambito Cybersecurity e IA

- **NORMATIVE PRINCIPALI:**
  - **GDPR** (Regolamento UE 2016/679)
    Fondamentale per la protezione dei dati personali e la sicurezza nel trattamento automatizzato.
    Prevede principi come minimizzazione, privacy by design e accountability.
  - **NIS 2** (Direttiva UE 2022/2555), entrata in vigore il 18/10/2024
    Rafforza gli obblighi di sicurezza per le infrastrutture critiche, aziende IT e servizi essenziali.
    Ha previsto aggiornamenti periodici e adempimenti con scadenza al 31/07/2025.
  - **Codice dell'Amministrazione Digitale** (Italia) - ultimo aggiornamento maggio 2024.
    Regola la sicurezza dei sistemi pubblici e la gestione dei dati nel settore pubblico italiano.
  - **EU AI Act** (Regolamento UE 2024/1689)
    La prima normativa orizzontale europea per la regolazione dell'IA.
    - classificazione del rischio; • trasparenza; • valutazioni d'impatto; • mantenimento della Governance; • sanzioni fino al 7 % del fatturato globale.
- **ASPETTI CRITICI IN AMBITO CYBER-AI:**
  - Responsabilità legale per danni causati da sistemi IA autonomi;
  - Obblighi di trasparenza e conformità (es. GDPR) per algoritmi che trattano dati personali;
  - Necessità di condurre AI impact assessment e garantire tracciabilità;
  - Tempistica ancora incerta per alcune norme: sono in corso consultazioni pubbliche e possibili aggiustamenti regolatori.

IA come strumento di ATTACCO

IA come strumento di ATTACCO 1/3

L'IA come forza moltiplicatrice per l'attacco

L'efficacia degli attacchi cresce con l'automazione e l'ottimizzazione offerte dall'IA, che aumenta la velocità e la precisione delle operazioni offensive:

- Vantaggio asimmetrico per gli attaccanti:
  alta specializzazione, scelta del momento e conoscenza del contesto.
- IA come amplificatore dello squilibrio:
  migliora tecniche esistenti, introduce nuovi paradigmi di attacco.
- Generative AI:
  phishing senza errori grammaticali, multilingua e personalizzato e difficile da rilevare.

Tabella: Percentuale di attacchi informatici influenzati dall'IA - Italia 2024

TRIMESTRE	1°	2°	3°	4°
% ATTACCHI	35	38	18	30

+167% rispetto al 2023

(Fonte: Rapporto CLUSIT 2025 - Dati sugli attacchi informatici in Italia nel 2024)

IA come strumento di ATTACCO 2/3

Evoluzione della Social Engineering con IA

La combinazione tra tecniche di social engineering e l'utilizzo dell'intelligenza Artificiale Generativa rappresenta una minaccia evoluta e facilmente scalabile.

- **Tecniche classiche potenziate**:
  - *Pretexting*: IA crea identità credibili e contestualizzate.
  - *Phishing/Vishing*: L'integrazione tra modelli linguistici di grandi dimensioni (LLM) e tecnologie di sintesi vocale consente di generare messaggi e voci estremamente realistiche, riproducibili su larga scala.
  - *Baiting*: IA genera esche digitali mirate e dinamiche.

- L'uso dell'Open - Source Intelligence **(OSINT)**, combinato con l'analisi di pattern comportamentali ricorrenti, crea un terreno fertile per lo sviluppo di soluzioni di automazione intelligente.

Dal phishing di massa allo spearfishing automatizzato:
IA produce migliaia di attacchi personalizzati, combinando **efficacia e scalabilità**.

IA come strumento di ATTACCO 3/3

Nuove frontiere dell'attacco: malware, agenti autonomi e contromisure

Tecniche emergenti basate su AI:

- **Malware polimorfo**: codice che muta grazie a modelli generativi → elusione dei sistemi di difesa.
- **Face Recognition Spoofing**: GANs generano "volti master" → aggiramento dei controlli biometrici.
- **Ransomware avanzati**: adattivi, selettivi e crittografia evoluta.

AI Agents autonomi: una nuova minaccia:

- Scoprono, sfruttano e mantengono l'accesso **senza intervento umano**, in tempo reale.
- Automatizzano l'intera kill chain, **imprevedibili e rapidi**.

Contromisure necessarie: **Difesa proattiva basata su AI**:

- Rilevamento e risposta autonomi
- Analisi comportamentale real-time
- Simulazioni automatiche (es. red teaming AI)

IA come strumento di DIFESA

IA come strumento di DIFESA 1/3

L'Intelligenza Artificiale come Strumento Proattivo di Cyber Difesa

Ruolo strategico dell'IA nella cybersecurity:

- Evoluzione degli attacchi informatici: più sofisticati e meno rilevabili.
- L'IA abilita un approccio proattivo: rilevamento, analisi e risposta in tempo reale.
- Machine Learning e analisi predittiva: anticipazione di vulnerabilità e adattamento dinamico.

Rilevamento e automazione:

- Analisi in tempo reale di grandi volumi di dati → identificazione di pattern anomali.
- Rapidità d'intervento: riduzione dell'impatto degli attacchi.
- Automazione delle contromisure: isolamento dispositivi, revoca credenziali e patching.

4 pilastri della sicurezza abilitati da IA:

- **Sicurezza dei dati**: database autonomi, replica distribuita e cifratura automatica.
- **Sicurezza applicativa**: generazione di codice sicuro, "low - code" IA - driven.
- **Sicurezza identità**: autenticazione biometrica avanzata e scalabile.
- **Sicurezza di rete**: Zero Trust Packet Routing (ZPR), percorsi autorizzati basati su IA.

IA come strumento di DIFESA 2/3

SOC e Automazione delle Risposte con Intelligenza Artificiale

Security Operations Center (SOC) potenziato:

- Funzione: monitoraggio e risposta centralizzata alle minacce cyber.
- IA nei SOC → espansione della capacità decisionale e rapidità di intervento.

Tecnologie chiave per il rilevamento:

- **EDR** (Endpoint Detection and Response)
- **NDR** (Network Detection and Response)
- **UEBA** (User and Entity Behavior Analytics)
- **SIEM** (Security Information and Event Management)

Automazione della risposta agli incidenti:

- IA supporta l'orchestrazione e risposta semi-automatica.
- Strumento chiave: **SOAR** (Security Orchestration, Automation and Response)
  - Coordinamento automatico dei flussi di risposta;
  - Maggiore efficienza del SOC;
  - Analisi veloce e suggerimenti per le contromisure.

IA come strumento di DIFESA 3/3

Difesa Passiva e Attiva basata su Intelligenza Artificiale

- **Difesa Passiva: prevenzione e rilevamento**
  - **IDS/IPS intelligenti**:
    - Rilevamento di attacchi noti e zero-day tramite ML.
    - Analisi continua del traffico.
  - **Analisi dei log**:
    - NLP per eventi testuali e clustering per pattern rari.
    - Esempio: SIEM IA-powered.
  - **UEBA**:
    - Profilazione dei comportamenti utente.
    - Alert in caso di attività anomale → accesso da IP insoliti e orari atipici.
- **Difesa Attiva: contenimento automatizzato**
  - **SOAR e Hyper SOC**:
    - Analisi, correlazione e risposta accelerata agli incidenti.
    - Suggerimenti automatici di azione → contenimento e neutralizzazione.
    - Azioni su larga scala senza intervento umano diretto.
  - **Sintesi**
    - L'IA è un abilitatore critico di cybersecurity moderna.
    - Integra rilevamento predittivo, automazione decisionale e risposta adattiva.
    - Riduce il carico umano e aumenta l'efficienza nella gestione delle minacce.

IA come SUPERFICIE di ATTACCO

IA come SUPERFICIE di ATTACCO 1/3

Minacce e Vettori

Contesto e motivazioni:

- L'IA è pervasiva (veicoli autonomi, filtri spam e riconoscimento facciale).
- L'adozione crescente aumenta la superficie di attacco.
- Il 73% delle aziende ha centinaia/migliaia di modelli in produzione.
- Il 41% delle aziende ha già riscontrato incidenti di sicurezza legati all'IA.
- A breve, ~30% degli attacchi informatici coinvolgerà l'IA.

Tipologie principali di attacchi:

1. **Adversarial Attacks (Attacchi Adversariali)**
   - Input leggermente modificati → errori nel modello.
   - Tecniche: *Fast Gradient Sign Method (FGSM)*.
   - Esempi reali:
     - Adesivi su segnali stradali → errore nei veicoli autonomi.
     - Occhiali "adversariali" → bypass biometrici (Face-ID).
     - Comandi vocali nascosti (ultrasuoni) → attivazioni non volute.
2. **Data Poisoning (Avvelenamento dei Dati):**
   - Manipolazione del training set → apprendimento distorto.
   - Obiettivi: degradare le performance o introdurre *backdoor*.
   - *Backdoor Attack*: inserimento di trigger invisibili che causano output desiderato dall'attaccante solo in casi specifici.

IA come SUPERFICIE di ATTACCO 2/3

Furto di Modelli e Attacchi alla Privacy

3. Model Stealing (Furto di Modelli):

- I modelli sono asset preziosi → economicamente e strategicamente.
- **Approccio Black-box**: input - output raccolti via API → clone del modello.
- **Approccio Side-channel**: segnali fisici (latenze, consumi) usati per inferire dettagli interni.
- Rischi associati:
  - Studio offline di vulnerabilità.
  - Re-ingegnerizzazione e iniezione di versioni modificate nei sistemi vittima.

4. Data Leakage (Fuga di Dati):

- I modelli possono “ricordare” dati del training → rischio privacy.
- **Membership Inference**
  - Determina se un dato era nel training set.
  - Indicatore: output a confidenza elevata su dati noti → sintomo di overfitting.
  - Esempio: pazienti identificati in un dataset medico.
- **Model Inversion**
  - Ricostruzione di dati di training partendo dagli output.
  - Esempio: rigenerazione di volti da modelli di riconoscimento facciale.

IA come SUPERFICIE di ATTACCO 3/3

Casi Studio, Implicazioni e Mitigazioni

Esempi pratici:

- Keen Security Lab: adesivi ingannano l'autopilota Tesla → deviazioni di corsia.
- Carnegie Mellon: occhiali "adversariali" → identificazioni false con successo elevato.
- Comandi ultrasonici: esecuzione di azioni invisibili per l'utente.

Caso DeepSeek:

- Dubbio sulla reale originalità del modello rispetto a GPT-4.
- Rischio di "memorization leakage" da dati pubblici generati da ChatGPT.

Mitigazioni possibili:

- Privacy differenziale: rumore statistico nei dati per ridurre memorizzazione puntuale.
- Regolarizzazione: limita l'overfitting → riduce leak di dati.
- Auditing Tools: valutano la sicurezza e la robustezza dei modelli AI, identificando punti deboli che potrebbero essere sfruttati da attaccanti.
- Test di *Membership Inference* interni valutazione del rischio di fuga informativa.
- Validazione continua di robustness e sicurezza dei modelli AI in produzione.

L’etica nella IA e Cybersecurity

Corso di Perfezionamento in Cybersecurity, Cyber Risk and Data Protection

L’etica nella IA e Cybersecurity

L’adozione dell’IA nella sicurezza informatica solleva questioni che non sono solo tecniche, ma anche morali e sociali.
L’etica diventa quindi un elemento strategico per orientare l’uso dell’IA verso finalità giuste, trasparenti e inclusive.

Sfide e interrogativi:

- Come garantire che un sistema IA non discrimini o amplifichi pregiudizi?
- È accettabile che un algoritmo, in nome della sicurezza, limiti la libertà o la privacy?
- Chi è responsabile di un errore commesso da un agente autonomo?

Verso un uso responsabile:

- **Supervisione umana** nei processi decisionali critici;
- **Trasparenza e auditabilità** dei modelli, specialmente quelli che incidono su persone e diritti;
- **Proporzionalità e finalità**: usare l’IA dove serve, non dove può;
- **Inclusività**: l’IA deve essere progettata per tutelare, non escludere.

Un’IA che ignora l’etica può diventare essa stessa una minaccia.
Integrare l’etica nella progettazione significa rafforzare sicurezza, fiducia e sostenibilità.

Conclusioni

Conclusioni

L’analisi condotta ha evidenziato come l’IA occupi un ruolo sempre più centrale nella Cybersecurity, in un delicato equilibrio tra **opportunità** e **minacce**.

Come **strumento di attacco**, l’IA consente operazioni sempre più sofisticate e automatizzate, potenziando in particolare le tecniche di social engineering e l’impiego di malware intelligenti.
Sul fronte **difensivo**, si afferma come tecnologia abilitante per il rilevamento proattivo delle minacce, l’analisi comportamentale e l’automazione delle risposte.
Parallelamente, l’IA si configura anche come **superficie di attacco**, esponendo nuovi vettori legati ai dati, ai modelli e all’opacità decisionale.

In un contesto di rapida evoluzione tecnologica, è essenziale adottare approcci integrati che considerino non solo dei fattori tecnici, ma anche le **implicazioni giuridiche** ed **etiche**, applicando principi come il **privacy by design** e l’**ethics by design**, così da garantire che questi aspetti siano incorporati fin dall’inizio nei processi progettuali.

In **conclusione**, la sfida attuale richiama l’attenzione sulla progettazione di sistemi di Intelligenza Artificiale che, oltre a essere sicuri per gli esseri umani, garantiscano anche la protezione dell’IA stessa e la prevenzione dei potenziali rischi derivanti dal suo utilizzo.
Una strada possibile è quella dell’**equilibrio**: un’IA responsabile, supervisionata dall’essere umano e sostenuta da un quadro normativo solido.