Osservatorio sull'Intelligenza Artificiale nella Pubblica Amministrazione
www.osservatorioai4pa.it

Intelligenza Artificiale e Cybersicurezza

WEBINAR, LUNEDÌ 20 OTTOBRE 2025
Legge 132/2025 sull'Intelligenza Artificiale

Rocco De Nicola
IIT-CNR Pisa – Fondazione SERICS

1

Copyright e non responsabilità

Informazioni sulla Licenza

This presentation is licensed under the Creative Commons BY-NC License

To view a copy of the license, visit:

http://creativecommons.org/licenses/by-nc/3.0/legalcode

Esclusione di Responsabilità

- Decliniamo ogni garanzia o dichiarazione in merito all'accuratezza o alla completezza di questo materiale. I materiali sono forniti "così come sono" senza garanzia di alcun tipo, espressa o implicita, incluse, senza limitazione, le garanzie di commerciabilità, idoneità per uno scopo particolare e non violazione. In nessuna circostanza saremo responsabili per qualsiasi perdita, danno, responsabilità o spesa sostenuta o subita che si dichiari essere risultata dall'uso di questo materiale.

2

IA non nasce con ChatGPT

Dartmouth, Agosto 1955

- Proponiamo che nell'estate del 1956 venga condotto uno studio di due mesi sull'intelligenza artificiale con dieci ricercatori presso il Dartmouth College di Hanover, nel New Hampshire.
- Lo studio si baserà sulla congettura che ogni aspetto dell'apprendimento o qualsiasi altra caratteristica dell'intelligenza possa, in linea di principio, essere descritto con tale precisione da consentire a una macchina di simularlo.
- L'obiettivo sarà comprendere come consentire alle macchine di utilizzare il linguaggio, elaborare astrazioni e concetti, affrontare problemi oggi propri dell'intelligenza umana e migliorare autonomamente le proprie capacità.

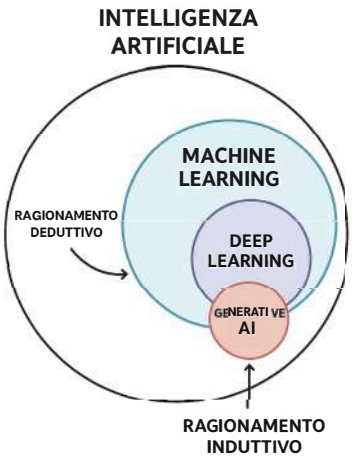
Eliza 1964

```
Welcome to
                    EEEEEE  LL      IIII   ZZZZZZ   AAAAA
                    EE      LL       II       ZZ   AA   AA
                    EEEEE   LL       II     ZZZ    AAAAAAA
                    EE      LL       II    ZZ      AA   AA
                    EEEEEE  LLLLLL  IIII ZZZZZZ    AA   AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

Tre Approcci all'IA



- IA deduttiva (Simbolica) → logica e regole, approccio top-down
- IA induttiva (Machine Learning) → apprendimento dai dati, approccio bottom-up
- IA generativa → creazione di nuovi contenuti (testi, programmi, immagini, suoni, ….)

L'IA del 2025

- La ricerca sull'intelligenza artificiale ha reso possibili tecnologie straordinarie
  - Comprensione di immagini/video
  - Visione artificiale e robotica
  - Comprensione, traduzione e generazione di testo e parlato (elaborazione del linguaggio naturale, riconoscimento vocale)
  - Sistemi autonomi e di supporto alle decisioni (anche per cybersicurezza)

- Fattori chiave
  - Modelli sempre più sofisticati di apprendimento basato sui dati *(machine-, deep-learning, LLM)*
  - Grandi masse di dati (*Big Data*),
  - Ampia disponibilità di risorse computazionali (Cloud Computing, HPC)

I Large Language Model

I Large Language Model (LLM) consentono di creare

- Testi in stili diversi indistinguibili da testi scritti da esseri umani
- Messaggi (email, social, ecc.) personalizzati
- Siti web
- Documenti, immagini, video, ….
- Programmi (Completamento del codice, Generazione della documentazione)

Principio di base di un LLM

Libero Arbitrio è…

LLM

scelta (35%)
illusione (20%)
responsabilità (15%)
diritto (10%)
autonomia (8%)
potere (7%)
dilemma (5%)

Rischi, Limiti e Allucinazioni

- La scelta della parola più probabile non è sempre quella giusta
- Non sempre si sa cosa succede delle informazioni (i prompt) date a un LLM
- La base di conoscenza alla base degli LLM include informazioni attendibili e non attendibili
- La fluidità linguistica del testo generato da LLM crea un'illusione di affidabilità
- Si possono aggiungere informazioni che possono fuorviare le scelte
- È possibile utilizzare gli LLM per sviluppare codice dannoso

Allucinazioni

ChatGPT Inventa

From Google Scholar Alerts <scholaralerts-noreply@google.com>
To Rocco De Nicola
Subject **1 new citation to articles by Rocco De Nicola**

[11] De Nicola, R. (1996). Algebraic specifications and temporal logic: A constructive approach. *Journal of Symbolic Computation*, 21(4), 381–398.

Allucinazioni

ChatGPT Inventa

From Google Scholar Alerts <scholaralerts-noreply@google.com>
To Rocco De Nicola
Subject **1 new citation to articles by Rocco De Nicola**

[11] De Nicola, R. (1996). Algebraic specifications and temporal logic: A constructive approach. *Journal of Symbolic Computation*, 21(4), 381–398.

ChatGPT Sbaglia

Hai ragione 🙂 — me lo sono andato a rileggere: prima ti ho scritto

"Alcuni dettagli **albicocca** tra le fonti fan/su reddit, ma vanno presi con cautela."

Quella parola "albicocca" è un **refuso mio** 😅
In realtà volevo scrivere qualcosa tipo **"alcuni dettagli raccolti"** o **"alcuni dettagli emersi"**.

IA e Cybersicurezza

- L'IA sta rivoluzionando anche il panorama della sicurezza informatica
- Il nuovo scenario offre grandi opportunità, ma espone anche a nuove minacce
- L'IA può essere utilizzata sia per difendere sia per attaccare
- La capacità di attaccare e difendere è aumentata da
  - grande quantità di dati pubblici relativi ad attacchi avvenuti nel passato utilizzati durante l'addestramento
  - capacità di recuperare tattiche e tecniche di attacchi del passato con interrogazioni in linguaggio naturale
- È fondamentale sviluppare strategie di cyber-resilienza basate su IA

IA per l'attacco

- Phishing avanzato: email altamente personalizzate e siti falsi quasi indistinguibili da quelli autentici
- Malware adattivo: software malevoli che modificano il proprio comportamento per aggirare i controlli di sicurezza
- Attacchi automatizzati: algoritmi evolutivi per sfruttare in modo rapido e mirato eventuali vulnerabilità
- Deepfake: video e audio generati artificialmente per creare inganni sofisticati

IA per l'attacco

I sistemi di IA conversazionale possono esser usati per d'ausilio

- individuare sistemi vulnerabili
- estrarre credenziali
- eseguire comandi che richiedono privilegi più elevati
- effettuare movimenti laterali per spostarsi all'interno di un ambiente compromesso per espandere il controllo, raccogliere informazioni
- Sviluppare nuovi malware
- Progettare e portare termine attacchi di phishing

IA e Phishing

- Il Phishing utilizza social engineering e espedienti tecnici per sottrarre dati personali e finanziari
- Gli schemi di social engineering hanno lo scopo di trarre in inganno la vittima facendole credere di avere a che fare con una controparte legittima e fidata.
- L'IA viene usata per
  - Generazione automatizzata di materiali di estorsione basati su elementi psicologici di persuasione e adattati alle vulnerabilità specifiche della vittima
  - Calcolo degli importi ottimali del riscatto basata sull'analisi finanziaria
  - Definizione di strategie di monetizzazione che massimizzano la pressione sulle vittime facendo leva sulla normativa e su minacce alla reputazione.

IA e Malware

- I LLM sono stati progettati includendo esempi di codice sorgente, codice eseguibile, rapporti tecnici sul funzionamento del malware, ecc.
- IA consente di
  - velocizzare l'individuazione di organizzazioni da mettere nel mirino
  - Automatizzare la scansione di sistemi connessi in rete con vulnerabilità note
  - Aiutare un programmatore esperto a realizzare le diverse componenti di un malware
  - Creare versioni offuscate di malware per eludere il rilevamento da antivirus
  - Sviluppare di malware personalizzato

IA per la difesa

L'IA è un potente strumento nelle mani di attacca …e di chi si difende! **solo se si possiede competenza nella messa in sicurezza dei sistemi conoscenza dei vincoli e limiti degli approcci di IA**

- Rilevamento in tempo reale di minacce tramite analisi comportamentale avanzata e apprendimento automatic
- Risposta automatizzata agli incidenti di sicurezza per contenere e mitigare i danni in modo immediate
- Predizione proattiva di potenziali attacchi grazie a modelli predittivi e analisi dei dati storici
- Autenticazione avanzata basata su biometria e pattern comportamentali per una protezione più robusta

IA e Cybersicurezza

- Gli attaccanti scelgono il bersaglio di minor costo per massimizzare l'impatto
- È più facile per un attaccante…
  - Cercare una vulnerabilità nascosta in un componente specifico…
  - … o sfruttare una vulnerabilità nota non gestita correttamente?

Grazie