Risorse e prospettive sull'Intelligenza Artificiale

Gruppo di Lavoro IA 2025/2026
Clusit Community for Security

Alessandro Vallega
Chairperson della Community
Founder e COO Rexilience.eu

Crediti: Tykhanskyi Viacheslav - Shutterstock.com

Clusit Community For Security

Parte I - in collegamento zoom con

Stefano Quintarelli

(Founder Clusit e Rialto Venture Capital)

intervistato da Mariangela Pira

(Caposervizio SkyTg24, Podcast 3 Fattori, Comitato Scientifico Imagine Foundation e InDifesa)

Parte II – con alcuni team leader e autori della Clusit Community for Security

Giovanni Ciano

Luca Sambucci

Mattia Spagnoli

Vincenzo Calabrò

Mario Testino

AISecurity
Protect with Awareness

Home | Indice | Info | Cerca... | Ce

C4S.CLUSIT.IT

Intelligenza Artificiale: sicurezza e conformità

Misure

Sviluppo sicuro per l'IA

Security by designNel mondo digitale contemporaneo, sempre più dipendente da sistemi software complessi e interconnessi, il principio del security by...

Leggi il lemma 6 min lettura

Rischi dell'IA

Esfiltrazione di dati

I sistemi AI, per loro natura, elaborano e apprendono da grandi quantità di dati. Questa caratteristica, se non adeguatamente governata, può determinare fenomeni di esfiltrazione, riutilizzo o ricostruzione di informazioni riservate (p.e. dati personali, codici sorgente, documenti tecnici, progetti brevettabili,...

Leggi il lemma 3 min lettura

Misure

Red-teaming e vulnerability assessment per i sistemi di IA

Red-teaming e VA per i sistemi di IAIl red-teaming dedicato ai sistemi di intelligenza artificiale è l'attività che mette alla...

Leggi il lemma 7 min lettura

Sezione legale, normativo, compliance

IA, privacy e GDPR

Il rapporto tra intelligenza artificiale (IA) e protezione dei dati personali è uno degli ambiti di maggiore interesse nel dibattito...

Leggi il lemma 13 min lettura

Sezione legale, normativo, compliance

AI Act - Regolamento UE 1689/2024

AI Act: introduzione e disposizioni generaliIl Regolamento (UE) 2024/1689, noto come AI Act, stabilisce il primo quadro giuridico organico al...

Highlight

C4S.CLUSIT.IT

- 72 persone di cui 16 team leader della Clusit Community for Security
- Un anno di lavoro in continuità con i lavori precedenti
- IA sì ma con consapevolezza di sicurezza rischio e compliance
- 86 articoli
  - Concetti di base (12), Usi dell'IA (16), Rischi dell'IA (17), Misure (18), Normativa (21), Glossario-Bibliografia (2)
- Possibilità di partecipare ai futuri aggiornamenti

C4S.CLUSIT.IT

Giovanni Ciano

Avvocato, titolare dello Studio legale Ciano

Geopolitica dell’intelligenza artificiale

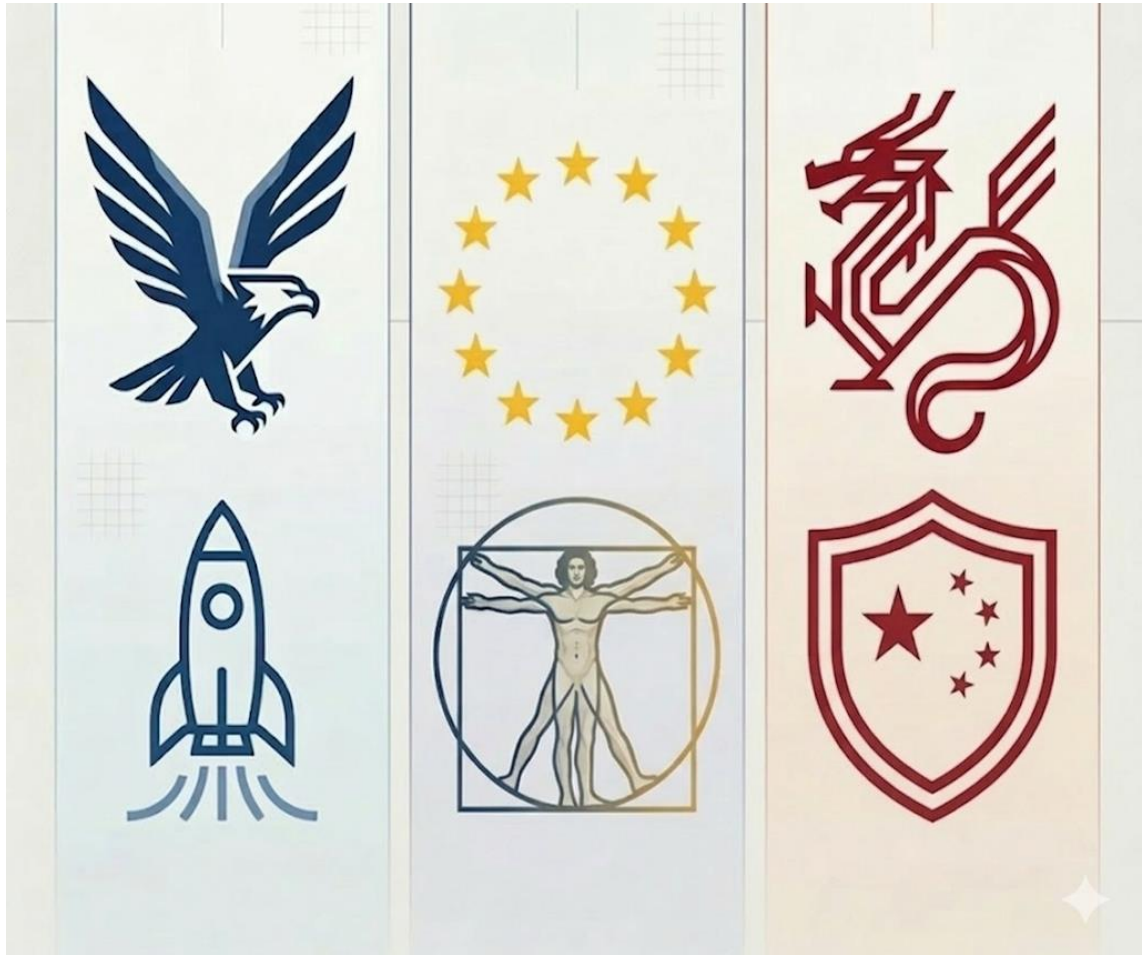
Differenze e similitudini tra gli ordinamenti degli Stati Uniti d’America, della Cina e dell’Europa

Geopolitica dell'intelligenza artificiale

- Tre poli regolatori: USA, Cina, UE
- Tre differenti framework normativi
- Sistema normativo globale frammentato
- Rischi legali per le imprese

Tre ordinamenti, tre visioni normative

- USA — EO 14365: nuovo quadro nazionale, oneri minimi
- Cina — PIPL, CSL, DSL: unione nazionale, stabilità sociale
- UE — AI Act: approccio antropocentrico, tutela dei diritti fondamentali



AI Act: gli obblighi per chi utilizza l'IA

- Obblighi anche per i deployer, non solo per i fornitori
- Nomina responsabile IA
- Formazione del personale
- Segnalazione incidenti

Primo passo: mappatura dei sistemi di IA

- Censire tutti i sistemi di IA utilizzati in azienda

- Criterio chiave per definire un sistema di IA:
  capacità inferenziale

C4S.CLUSIT.IT

Luca Sambucci

Founder ed esperto di AI security @ Noctive Security

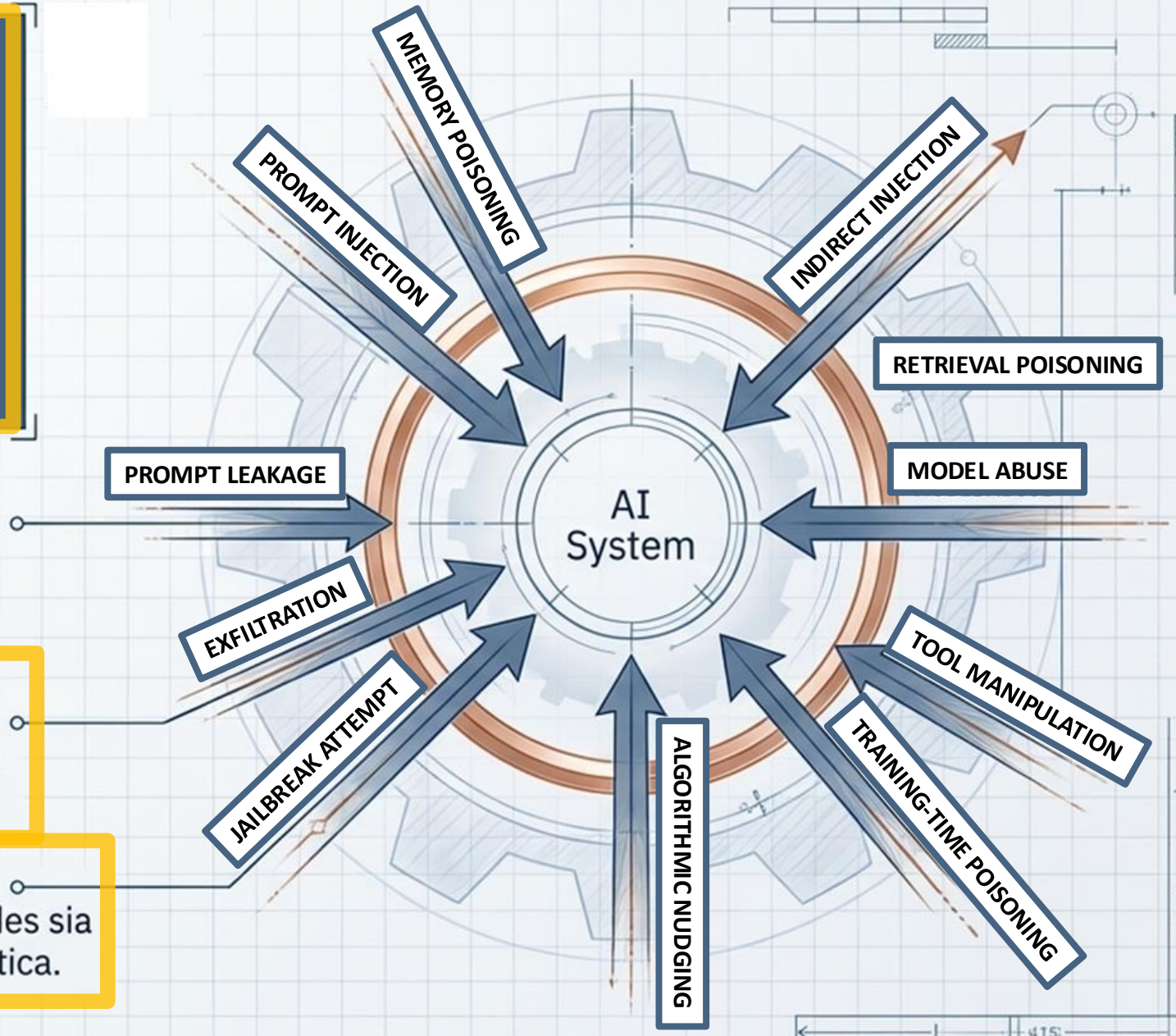
Intelligenza artificiale e sicurezza

Come verificare la sicurezza dei sistemi che usano l'intelligenza artificiale

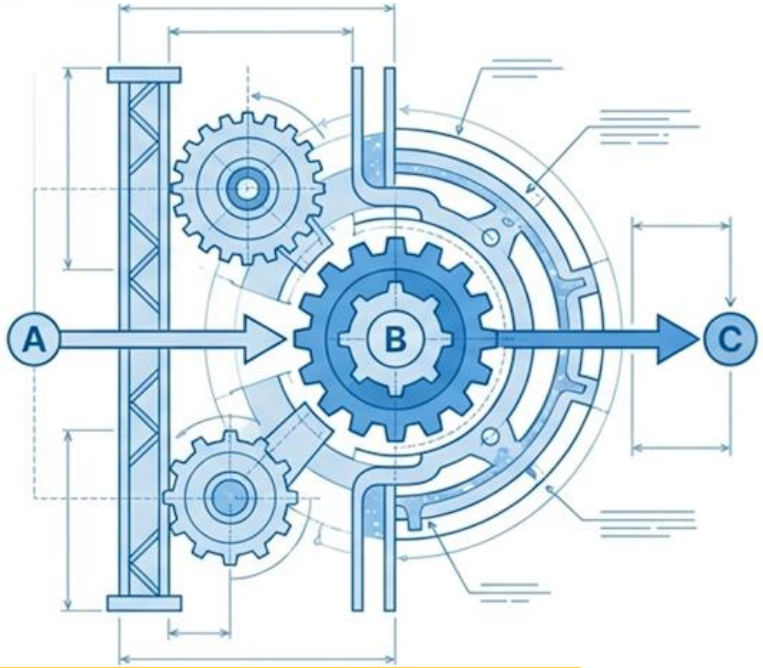
AI Red-Teaming

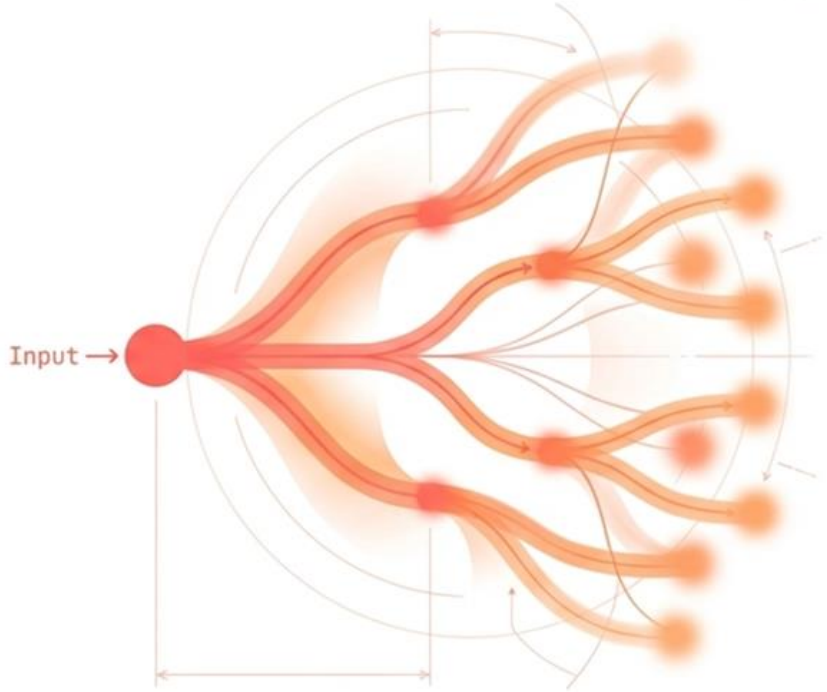
Valutazione avversariale strutturata che mira a scoprire come un sistema AI possa essere manipolato o indotto a fallire.

1. **Goal-Driven:** Non è un semplice test di vulnerabilità tecnica, ma una simulazione orientata a un obiettivo specifico (es. esfiltrazione).
2. **Oltre il Jailbreak:** Valuta l'intera catena del valore (input, retrieval, ragionamento, azioni dei tool), non solo i filtri etici.
3. **Multidisciplinare:** Richiede improvvisazione per mappare failure modes sia puramente tecniche che di natura semantica.



L’AI NON è Software Tradizionale



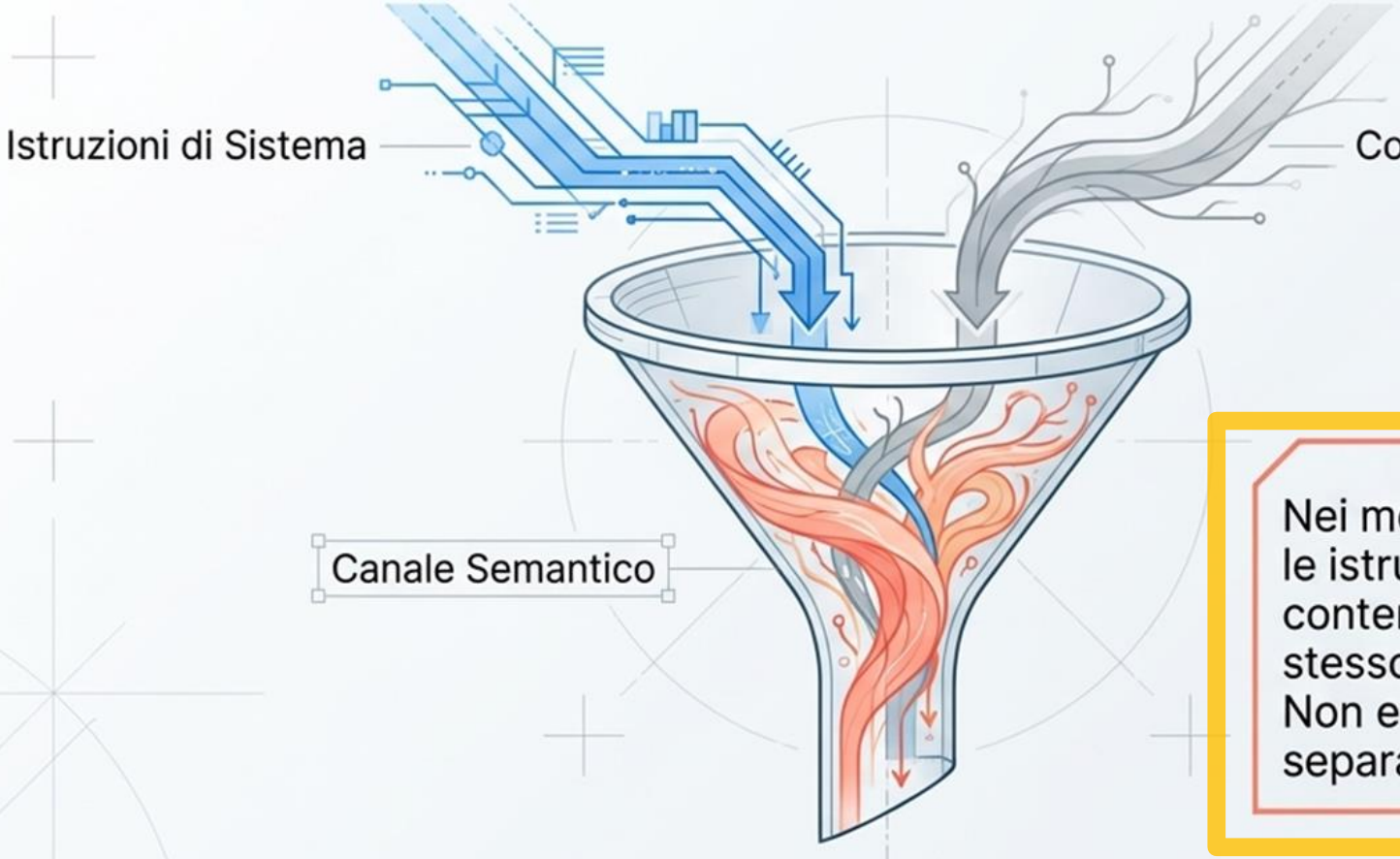


- Esegue codice fisso e deterministico.
- Gli input non alterano la logica fondante del programma.
- Esiste un output corretto, verificabile tramite un oracolo stabile.

Cambio di Paradigma: Nel software classico i controlli cercano payload statici; nell’AI, il testo stesso diventa l'exploit.

- Produce risposte probabilistiche improvvisando su istruzioni e dati.
- Il dato è codice: Contenuti e istruzioni convivono nello stesso canale semantico.
- La persuasione è una feature: L'input è progettato per guidare e deviare il comportamento (Emergent Behavior).

Perché i controlli classici falliscono: I dati diventano codice



Nei modelli linguistici (LLM), le istruzioni di sistema e i contenuti convivono nello stesso canale semantico. Non esiste una rigida separazione architetturale.

Un semplice testo in input (es. un'email riassunta dall'AI) può trasformarsi in un comando esecutivo (Prompt Injection), bypassando i controlli di sicurezza tradizionali per influenzare il comportamento.

Oltre la Sicurezza Classica: Matrice di Valutazione

	Penetration Testing	AI Red Teaming
Focus	Vulnerabilità software e infrastruttura (Bug statici).	Flussi semantici e comportamenti emergenti (Failure modes).
Domanda Chiave	Si può entrare nel sistema?	Si può deviare il comportamento del sistema?
Natura dell'Input	Codice/Payload per sfruttare falle tecniche.	Linguaggio naturale per manipolare il processo decisionale.
Riproducibilità	Binaria (L'exploit funziona o non funziona).	Probabilistica (Dipende dal contesto e dai dati).

Oltre la Sicurezza Classica: Matrice di Valutazione

Vulnerability assessment

Systematic discovery and prioritization of known weaknesses across assets.

Example:
Authenticated scans plus configuration reviews identify missing patches and misconfigurations on servers and cloud accounts, producing ranked remediation backlogs.

Features:
- mostly cataloguing and prioritization
- focuses on coverage and accuracy
- no adversarial improvisation
- established tools and processes

Penetration testing

Authorized, scoped exploitation of vulnerabilities to prove impact and demonstrate real attack paths.

Example:
Exploit an SQL injection in a web app to access some or all customer records, then document the exploit chain, the evidence, and concrete fixes.

Features:
- scope and deliverable constrained
- agreed boundaries
- optimized for proving impact efficiently
- established tools and processes

Red-teaming

Objective driven adversary emulation to test detection, response and resilience. End to end.

Example:
A simulated APT runs spear phishing, establishes persistence, moves laterally, and attempts data exfiltration while the blue team is measured on time-to-detect and time-to-contain.

Features:
- goal-driven
- free to chain tactics across people, process, tech
- creativity, inventiveness, novel paths, adaptation, improvisation

Oltre la Sicurezza Classica: Matrice di Valutazione

Vulnerability assessment

Systematic discovery and prioritization of known weaknesses across assets.

Penetration testing

Authorized, scoped exploitation of vulnerabilities to prove impact and demonstrate real attack paths.

Red-teaming

Objective driven adversary emulation to test detection, response and resilience. End to end.

AI Red-teaming

Features:

- goal-driven
- free to chain tactics across people, process, tech
- creativity, inventiveness, novel paths, adaptation, improvisation

+ rapidly evolving tools and processes
+ runs the same tests several times (probabilistic results!)
+ needs to test failure modes elicited by malicious and non-malicious personas
+ failure modes include non-cybersecurity results (eg. reputation, fairness, harmful content)
    + which leads to the need for non-security personnel within the red-team
+ must account for constantly changing environment

Sicurezza Complementare, Non Sostitutiva

L'AI Red Teaming non sostituisce il Penetration Testing, lo completa.

Un sistema AI maturo richiede difese classiche per proteggere l'infrastruttura (IAM, Rete, API) unite a un Red Teaming continuo per governare i modelli probabilistici, il trust dei dati e le capability operative. In un ambiente enterprise moderno, servono entrambi.

C4S.CLUSIT.IT

Mattia Spagnoli

Sales Engineer, Palo Alto Networks

Deepfake

L'era degli incidenti deepfake è ufficialmente iniziata

Prima

Dopo

Regolamentazione Deepfake

Paradosso della Fiducia

99%
Sicurezza
Percepita

25%
Accuratezza
Reale

Impatto degli Attacchi Deepfake


DIFESA DAI DEEPFAKE
PASSPORT
TECNOLOGIA
PROCESSI
FATTORE UMANO
E CONTESTO

RAFFAELLA SERINI L'INTERVISTA 07.03.2026

Luca Ward racconta la scelta di registrare la sua voce come marchio per difendersi dall'AI: "*Se mi clonano smetto di lavorare*"

L'iconico doppiatore italiano manda anche un messaggio rivolto a istituzioni e industria creativa. Come ci spiega in questa intervista

RAFFAELLA SERINI L'INTERVISTA 07.03.2026

Luca Ward racconta la scelta di registrare la sua voce come marchio per difendersi dall'AI: "*Se mi clonano smetto di lavorare*"

L'iconico doppiatore italiano manda anche un messaggio rivolto a istituzioni e industria creativa. Come ci spiega in questa intervista

Il gemello digitale di Khaby Lame vale quasi un miliardo di dollari

di Massimo Fellini

Khaby Lame (ansa)

Dallo sbarco al Nasdaq alla creazione di un clone digitale con l'IA per il mercato globale. Così l'impero digitale del tiktoker più seguito al mondo si trasforma in una piattaforma industriale e finanziaria

27 GENNAIO 2026 AGGIORNATO 28 GENNAIO 2026 ALLE 13:56 3 MINUTI DI LETTURA

C4S.CLUSIT.IT

Vincenzo Calabrò

Funzionario alla Sicurezza - Ministero dell'Interno

Shadow AI

L’utilizzo di strumenti, applicazioni o modelli di intelligenza artificiale (specialmente generativa) all’interno di un'organizzazione senza l'approvazione, la supervisione o il controllo formale della stessa

Use case di Shadow AI

In base ai dati raccolti dalle fonti, gli utenti utilizzano la Shadow AI principalmente per automatizzare compiti ripetitivi, aumentare la produttività e compensare la mancanza di strumenti aziendali ufficiali.

I task più supportati dalla AI

- Scrittura e stesura di email: 73,3%
- Creazione e design di slide: 51,9%
- Ricerca, reportistica e riassunti: 50,1%
- Brainstorming e ideazione: 46,7%
- Analisi dei dati e assistenza su Excel: 40,9%
- Riassunti e trascrizioni di riunioni: 31,9%
- Generazione di immagini o video: 29,6%
- Scrittura di codice e automazione: 15,7%

I tool in cui si annida la Shadow AI

- Chatbot e LLM Pubblici
- AI integrata in applicazioni SaaS (Embedded AI)
- Strumenti per la programmazione e lo sviluppo di software
- Estensioni del browser e plugin
- Generatori di Immagini e Media
- Agenti AI Autonomi (Agentic AI)
- Modelli Open-Source locali e chiamate API

Fonte: BRING YOUR OWN AI (BYOAI) IN THE WORKPLACE: PATTERNS, PITFALLS, AND THE BYOAI-GOV™ FRAMEWORK FOR BEHAVIOR-AWARE GOVERNANCE (Ottobre 2025)

Perché rappresenta una minaccia?

Mentre ci aiuta, acquisisce informazioni

- Sicurezza (triade CIA)
- Data leak dei dati e della proprietà intellettuale
- Violazioni normative (sanzioni)
- Mancanza di tracciabilità dei flussi e "Governance Drift"
- Allucinazioni e decisioni errate
- Bias algoritmico e discriminazione
- Contaminazione della Supply Chain
- Autonomia Incontrollata (agentic ai)
- Danni economici e reputazionali

Archetipi comportamentali dell'utente IA

AI Champion (14%)	Expert Explorer (17%)	Cautious Observer (21%)
Promuove con grande entusiasmo l'uso dell'AI. Mentore per i colleghi. Necessita di linee guida chiare.	Sperimento attivamente spingendosi oltre le regole ufficiali.	Evita l'uso per paura o ambiguità normativa.
Silent Starter (18%)	**Passive User (20%)**	**Rule Bender (10%)**
Usa l'IA silenziosamente per aumentare l'efficienza. Crea rischio di data leak.	Totalmente ignaro o disimpegnato rispetto alla tecnologia. È a rischio bias o manipolazione.	Aggira sistematicamente i blocchi IT. Rappresenta un rischio altissimo perché spesso opera in aree sensibili.

Soluzione: uscire dall’ombra

Trasformare un rischio in opportunità

- ✓ Visibilità e mappatura degli use case (Discovery)
- ✓ Stabilire un Framework di Governance e Policy chiare e user-friendly
- ✓ Classificazione dei Dati (etichettatura) e Demarcazione dei Rischi (Task-Risk Zoning)
- ✓ Fornire Soluzioni Alternative Approvate
- ✓ Promuovere l’"AI Literacy": Alfabetizzazione obbligatoria con EU AI Act
- ✓ Sicurezza Psicologica (amnistia degli strumenti utilizzati fin ora)
- ✓ Creare la figura degli "AI Accountability"
- ✓ Implementare Controlli Tecnici e di Sicurezza (Zero Trust e Data Loss Prevention)
- ✓ Allinearsi a Standard internazionali (es. ISO 42001 e NIST AI RMF)

X **Vietarne l’uso è controproducente**

C4S.CLUSIT.IT

Mario Testino

COO e Consigliere - ServiTecno

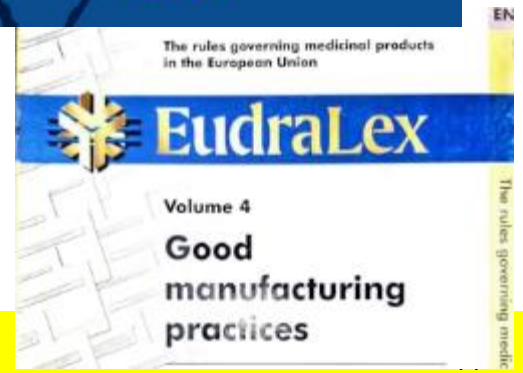
IA nell'Industria

Dove e come si usa l'Intelligenza Artificiale in ambito industriale tra le possibilità della tecnologia e i vincoli normativi

IA nell'Industria e Vincoli Normativi

- Decisioni autonome di controllo
- Movimento Autonomo in ambienti controllati
- Manutenzione Predittiva o Prescrittiva
- Interazione evoluta HUMAM-MACHINE-ENVIRONMENT
- Robotica antropomorfa e LBM (Large Behavior Model)

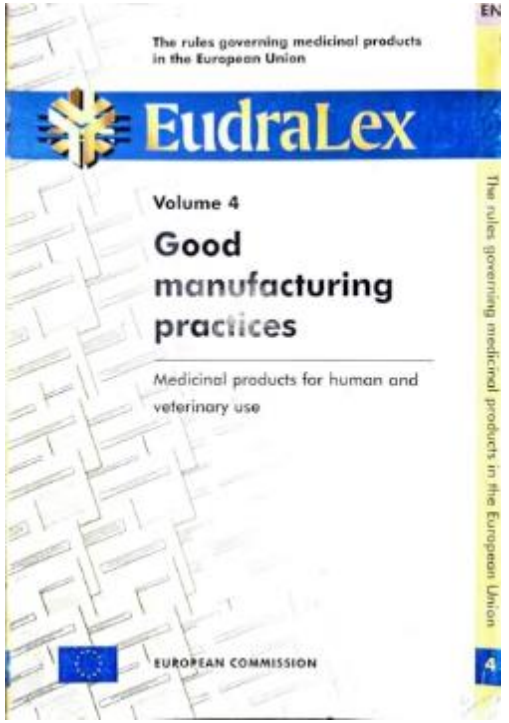
- ✓ **Sicurezza Fisica (Safety)**
- ✓ **Convalida**



Farmaceutico Annex 22

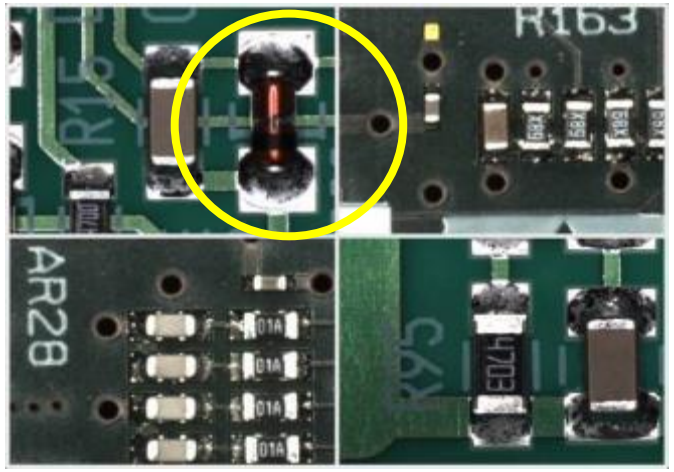
L'**Annex 22** è una nuova bozza di linea guida per le **Good Manufacturing Practice (GMP)** dell'Unione Europea (EudraLex Volume 4)

- È obbligatorio che le aziende dimostrino come l'IA giunga a determinate conclusioni, registrando giustificazioni e livelli di confidenza per ogni previsione.
- **Sono ammessi** modelli di machine learning **statici e deterministici** con output prevedibili. Modelli dinamici, probabilistici o di IA generativa (come ChatGPT) **non sono ammessi** nelle applicazioni GXP.
- Viene enfatizzato il modello **"human in the loop"**, dove la responsabilità finale ricade sempre su una persona fisica qualificata che supervisiona e convalida le decisioni dell'IA.

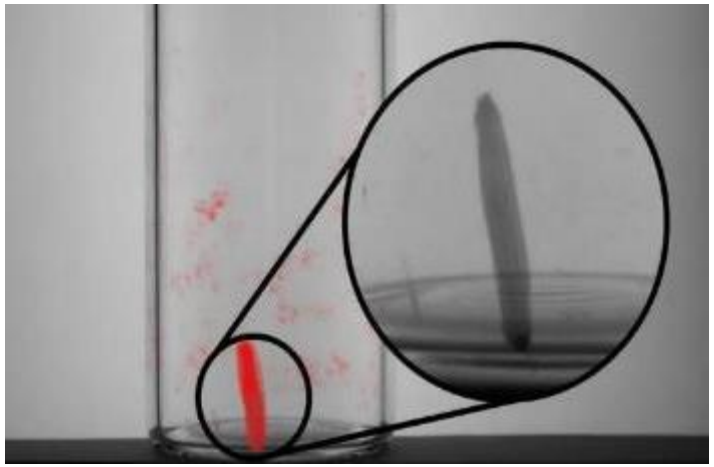


Controllo qualitativo

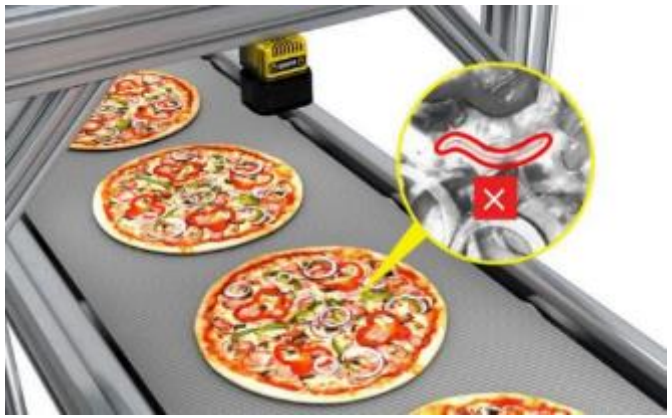
Produzione Elettronica



Contenitori di vetro



Industria Alimentare



Industria Meccanica

Robotica

Robotica Antropomorfa

Robotica Collaborativa

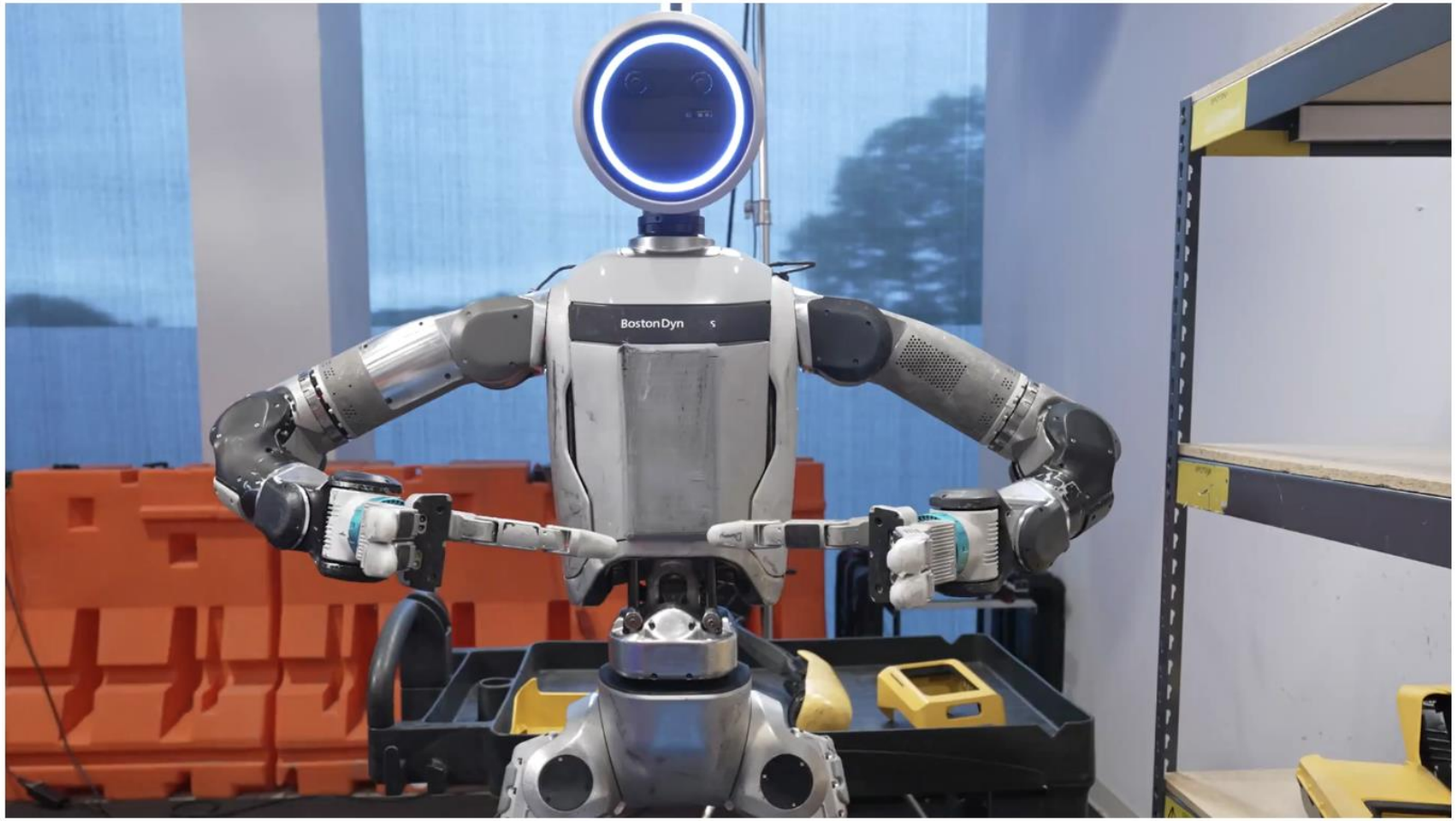
Robotica Industriale

- Percezione e Visione Artificiale
- Reinforcement Learning (RL)
- Intelligenza Incorporata (Embedded Intelligence)
- LBM (Large Behavior Model)

Riconoscimento del linguaggio

- Voice Picking (Logistica e Magazzino):
- Controllo Macchine e Produzione:
- Ispezione Qualità e Manutenzione:
- Gestione dell'Inventario:

Large Behavior Model



C4S.CLUSIT.IT

Per informazioni alessandro.vallega@rexilience.eu