Rapporto OAD 2025

Sistemi intelligenti a rischio: la mappa degli attacchi IA in Italia

Home

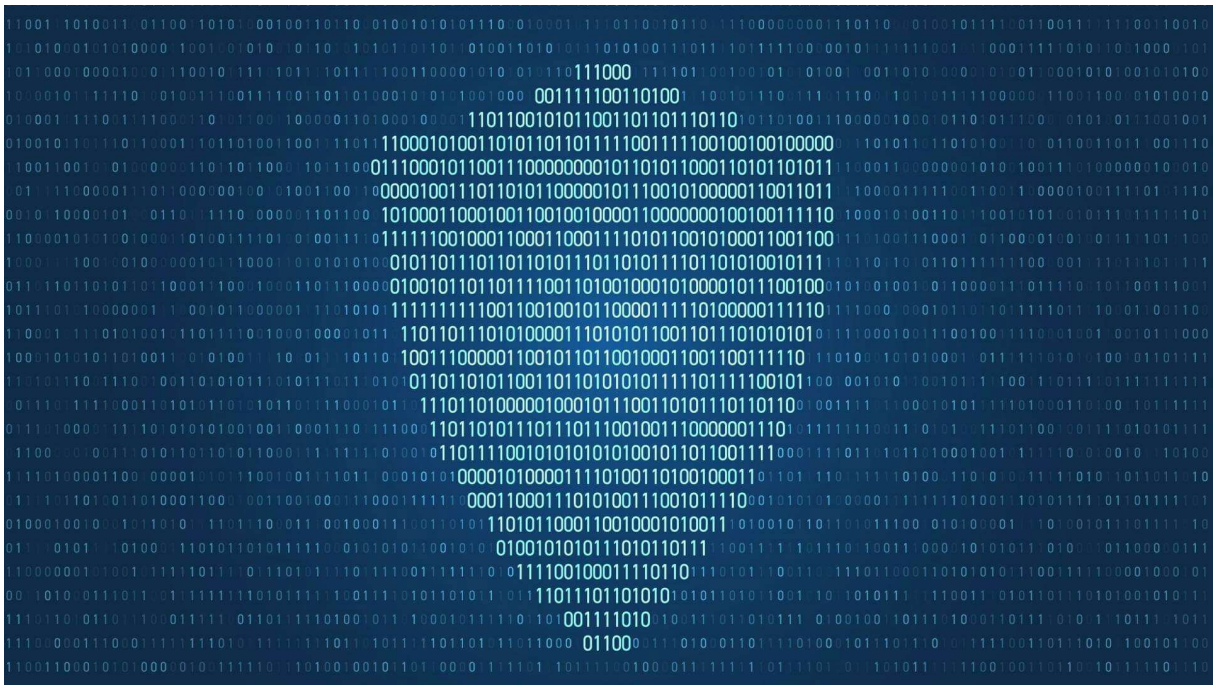
Sicurezza digitale

Il Rapporto OAD 2025 fotografa gli attacchi digitali in Italia, con focus su sistemi IA. Dati, settori colpiti, tecniche e impatti sono confrontati con Wef, Enisa, ACN e Polizia Postale. Evidenziate vulnerabilità Owasp LLM e raccomandazioni di resilienza operativa concreta

Pubblicato il 10 nov 2025

Marco R. A. Bozzetti

Presidente Onorario AIPSI – Associazione Italiana Professionisti Sicurezza Informatica



Negli ultimi dodici mesi gli **attacchi digitali in Italia** hanno mostrato dinamiche e impatti in linea con i principali osservatori internazionali, ma con tratti specifici legati al tessuto produttivo nazionale. A partire dai dati OAD 2025, ricostruiamo lo scenario, gli obiettivi e le priorità d'intervento, introducendo il focus sugli ambienti con Intelligenza Artificiale.

Cyber attacco costante alle aziende italiane: i dati OAD

Indice degli argomenti

Panorama 2024: profilo dei rispondenti e frequenza degli attacchi

L'indagine OAD 2025 prende in esame **gli attacchi intenzionali avvenuti nel 2024 ai Sistemi Informativi di aziende ed enti operanti in Italia,** ed approfondisce verticalmente gli attacchi agli ambienti web, a quelli OT (Operational Technology) e agli ambiti che utilizzano tecniche di Intelligenza Artificiale (IA).

Al questionario online OAD 2025, assolutamente anonimo, hanno risposto varie centinaia di interlocutori, soprattutto aziende dell'offerta ICT con un **21,4%** sul totale dei rispondenti (dai provider di servizi ICT ai consulenti ICT, dallo sviluppo software ai media digitali e alle telecomunicazioni). Seguono le industrie manifatturiere e di costruzioni, con un **14,85%**, la sanità e i servizi sociali con il **14, 41%**, i servizi professionali e di supporto alle imprese con un **10,48%**. Tutti gli altri raggruppamenti di settori merceologici, incluse le Pubbliche Amministrazioni Centrali e Locali hanno percentuali a decrescere dal 6% in giù.

Metodologia e campione dell'indagine OAD 2025

Il Rapporto OAD 2025[1] di AIPSI[2] in questa edizione (che rappresenta il diciottesimo anno consecutivo di indagini) per la prima volta ha aggiunto **l'Intelligenza Artificiale (IA) sia come target di attacchi (per sistemi/servizi applicativi basati su queste tecnologie) sia come strumento di attacco.**

OAD costituisce l'unica indagine indipendente online (via web) in Italia sugli attacchi digitali intenzionali ai sistemi informativi delle aziende e degli enti pubblici operanti in Italia, e sulle misure tecniche ed organizzative che questi hanno implementato. OAD non predefinisce uno specifico bacino di rispondenti, il medesimo negli anni, ma consente a chiunque, interessato e coinvolto nella gestione di un sistema informativo di una azienda/ente, un pieno e libero accesso al questionario online, in maniera totalmente anonima. Dato il numero di risposte raccolte e la loro distribuzione tra aziende ed enti pubblici di varie dimensioni e appartenenti a diversi settori merceologici, **l'indagine OAD fornisce preziose indicazioni sul fenomeno degli attacchi digitali intenzionali in Italia e delle misure di sicurezza in essere nei sistemi informativi delle**

imprese rispondenti. OAD riesce a coinvolgere nell'indagine anche le piccole e piccolissime realtà, che costituiscono in Italia la stragrande maggioranza e che le altre indagini nazionali ed internazionali difficilmente considerano ed analizzano.

Il presente articolo illustra il quadro generale emerso dall'indagine, inquadrandolo con quanto indicato dal World Economic Forum, da ENISA, da ACN e dal Servizio Polizia Postale e per la Sicurezza Cibernetica (che ha forniti dati e testi sul computer crime in Italia, si veda Cap.8 del Rapporto OAD 2025), ed approfondisce il tema degli attacchi digitali ai sistemi con IA.

In termini di dimensioni delle imprese rispondenti, come numero di dipendenti, hanno risposto all'indagine il **62,96%** di **piccole e medie organizzazioni**, ossia quelle fino a 250 dipendenti (le PMI, Piccole Medie Imprese, in ambito aziende private). Significativa, con un **22,22%**, la presenza tra queste delle **piccolissime organizzazioni con meno di 10 dipendenti**, che costituiscono la stragrande maggioranza delle imprese, private e pubbliche, in Italia (più del 95% del totale).

Il quadro nazionale degli attacchi digitali nel 2024

Questo è un elemento fondamentale e caratterizzante l'indagine OAD: nei suoi 18 anni consecutivi di attività, OAD è sempre riuscito a raccogliere dati sulla sicurezza digitale anche dalle organizzazioni medio-piccole, superando in ogni edizione il 60% del totale.

Il **71% delle organizzazioni rispondenti ha subito attacchi digitali nel 2024**.

La fig. 1-1 mostra la distribuzione percentuale degli **attacchi NON subiti** nel 2024 per dimensione (numero di dipendenti) delle aziende/enti rispondenti. La figura evidenzia come le aziende/enti rispondenti al di sotto dei 250 dipendenti rappresentano il 63% del totale, a conferma di quanto indicato sopra.

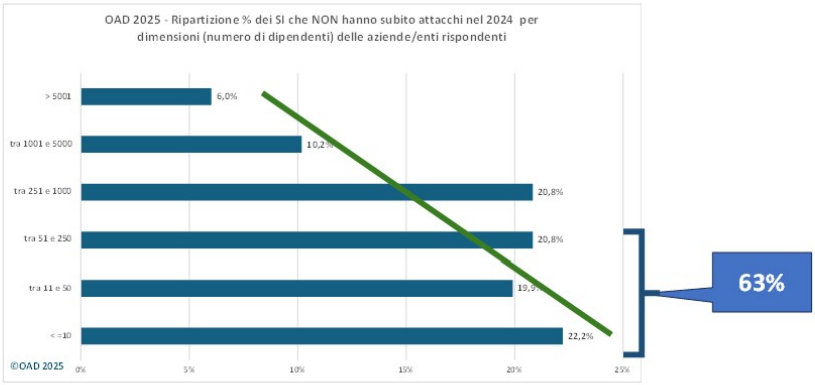


Fig. 1-1

La linea verde nella figura evidenzia come le organizzazioni dimensionalmente **più piccole** sono quelle che hanno subito/rilevato **meno attacchi digitali**. Esse infatti non rappresentano un obiettivo di interesse specifico per i cyber criminali, soprattutto per gli attacchi mirati (targeted attack), mentre possono essere coinvolte in attacchi di massa, come quelli basati sul phishing e sul ransomware, così come avviene tipicamente o per cogliere qualcuno nella massa, o per azioni di attivismo, di terrorismo informatico, di guerra digitale..

i **tipi di attacchi più diffusi** ("che cosa si attacca" tra le 16 famiglie di attacchi considerate nel questionario OAD 2025) includono:

- le **modifiche malevoli/non autorizzate ai programmi e alle configurazioni dei sistemi ICT,** con il **29,26%** delle risposte; a questo primo posto sicuramente contribuisce la larghissima diffusione di malware e di ransomware in Italia;
- l'**uso non autorizzato e malevolo di sistemi ICT del SI nel loro complesso**, con il **23,58%**;
- gli attacchi **DoS/DDoS**, per la saturazione dei sistemi ICT connessi ad Internet, con il **20%**: sono stati uno degli attacchi più usati nell'ambito delle cyber warfare, in particolare da parte di attaccanti pro Russia.

Da segnalare che gli attacchi **alla supply chain informatizzata,** considerata dal World Economic Forum uno dei principali rischi a livello mondiale, hanno avuto una diffusione tra i rispondenti del **14,85%**.

Tecniche impiegate e impatti operativi

Come tecniche di attacco ("come si attacca", considerando le 8 famiglie di tecniche del questionario OAD 2025) usate per il più grave attacco digitale subito, emergono le seguenti:

- con il **37,12%** l'uso di **più tecniche** di attacco contemporaneamente/in parallelo;
- con il **15,72%** la **raccolta illegale di informazioni**, tra cui il **social engineering**;
- con il **10,04%,** i **codici maligni**, alla base degli assai diffusi attacchi di ransomware.

Gli **impatti** derivanti dall'attacco più grave subito **sono significativi** in termini di disservizio, ed in molti casi con costi sia a livello di sistema informativo che di bilancio per l'intera azienda/ente:

- per i sistemi web, **64,3%** dei rispondenti ha subito interruzioni del servizio **≥2 giorni;**
- per i sistemi OT, per **> 85%** il blocco del sistema è durato **più di 2 giorni** e per il **30%** questo ha implicato un disservizio sul sistema informativo cui sono collegati per **2-3 giorni.**

Diffusione dell'intelligenza artificiale nelle imprese italiane e principali vulnerabilità per i sistemi LLM/GenAI

La fig. 2-1 mostra che circa il **41%** delle aziende/enti rispondential questionario OAD 2025 **utilizza** sistemi ed applicazioni basati su IA. Una % alta se si considera l'indagine europea DESI[3], che, come indicato in fig. 2-2, evidenzia che le imprese italiane che utilizzano qualsiasi tipo di tecnologia IA sono, nel 2024, **l'8,2%** del totale di imprese pubbliche e private con più di 10 dipendenti. Questo dato indica che le aziende/enti rispondenti si collocano in una fascia medio-alta per l'uso di strumenti informatici, oltre che per il livello di sicurezza digitale in essere nei loro sistemi informatvi.

Del 41% di rispondenti che utilizzano sistemi IA, il **23,8% ha rilevato attacchi**.

Analogamente a quanto fatto per gli attacchi ai sistemi web, per gli attacchi digitali ai sistemi di IA il questionario OAD 2025 ha fatto riferimento alle **10 principali e più diffuse vulnerabilità individuate a livello mondiale da OWASP[4]** per i sistemi IA di tipo LLM (Large Language Model) e di Generative AI**[5]**, che sono le applicazioni attualmente più diffuse. La tabella in fig. 2-3 elenca e brevemente illustra queste "top ten" vulnerabilità. La maggior parte riguarda i "prompt", ossia le informazioni che si passano al sistema di IA per avere le adeguate risposte, e le modalità di "istruzione" e messa a punto (di contenuti) del sistema stesso.

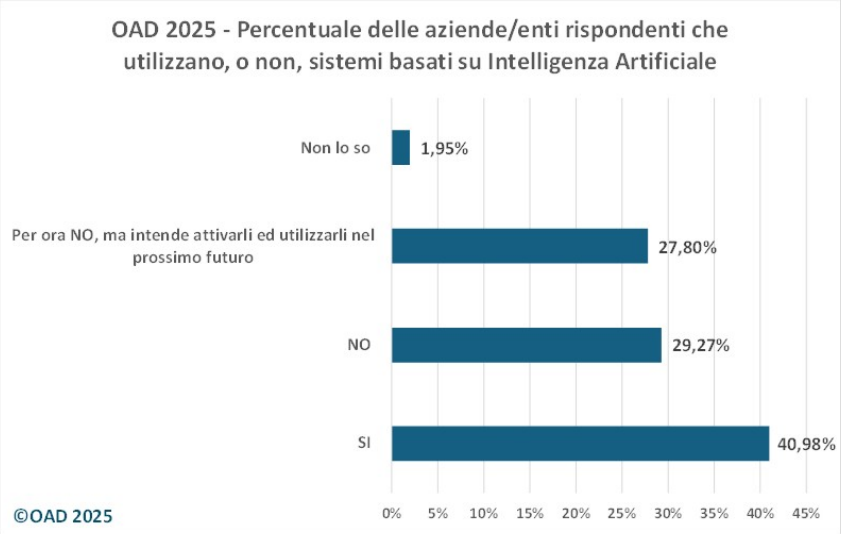


Fig. 2-1

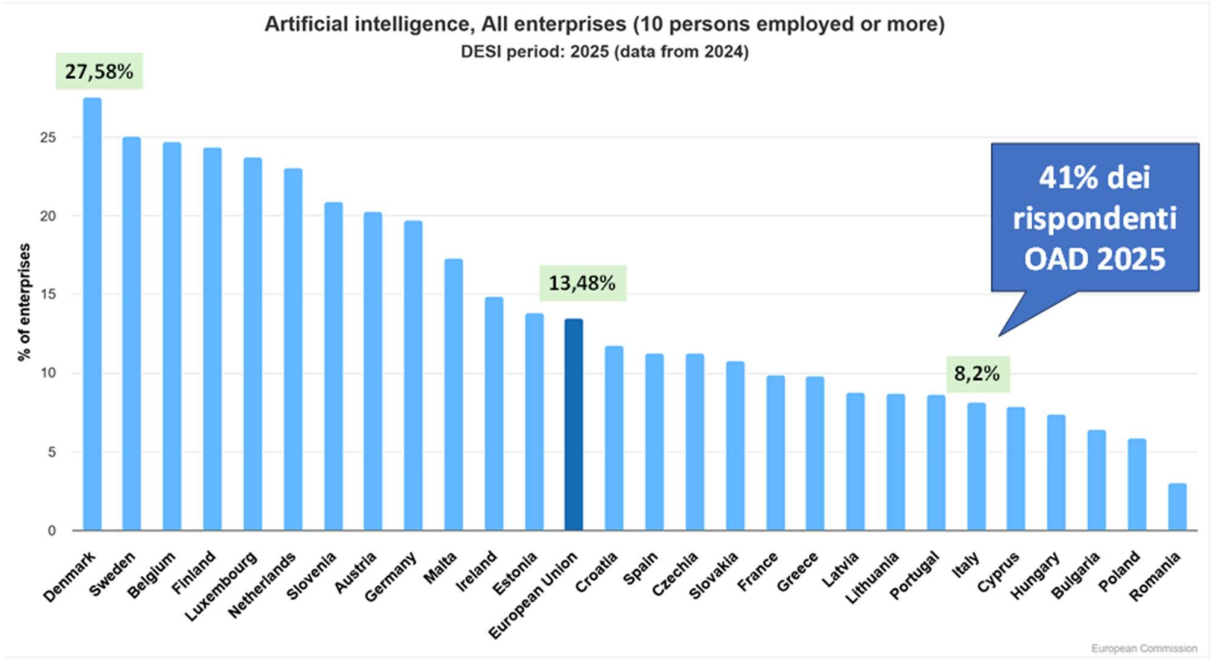


Fig. 2-2

Sfruttamento delle vulnerabilità e prime letture del fenomeno

La fig. 2-4 mostra, con risposte multiple, la **diffusione** percentuale tra i rispondenti a OAD 2025 dello sfruttamento (probabile) delle **Top Ten vulnerabilità elencate da OWASP** nell'attacco più grave ai sistemi di IA .

La vulnerabilità in testa a questa classifica per il bacino di rispondenti 2025 che hanno rilevato attacchi ai sistemi IA di tipo LLM e Gen-AI del loro sistema informativo, è il "**Data and Model Poisoning**"

(Avvelenamento dei dati e del modello) con un **35%** cui segue, con **20%**, l'"**Improper Output Handling**" (Gestione impropria dell'output).

Nome vulnerabilità	Breve descrizione
Prompt Injection (iniezione di prompt)	Questa vulnerabilità si verifica quando i prompt dell'utente alterano il comportamento o l'output dell'LLM in modi non intenzionali. Questi input possono influenzare il modello anche se sono impercettibili agli esseri umani, quindi le iniezioni di prompt non devono essere visibili/leggibili agli esseri umani, purché il contenuto venga analizzato dal modello. Le vulnerabilità di Prompt Injection sono presenti nel modo in cui i modelli elaborano i prompt e nel modo in cui l'input può forzare il modello a passare in modo errato i dati dei prompt ad altre parti del modello, causando potenzialmente la violazione delle linee guida, la generazione di contenuti dannosi, l'abilitazione di accessi non autorizzati o l'influenza di decisioni critiche.
Sensitive Information Disclosure (Divulgazione di informazioni sensibili)	Nella dizione statunitense i dati sensibili non sono solo quelli che fanno riferimento alle informazioni personali, in EU protette secondo i dettami della norma GDPR, ma tutte le informazioni "critiche" e confidenziali. Esse possono influenzare sia l'LLM che il suo contesto applicativo. Ciò include informazioni personali identificabili (PII), dettagli finanziari, cartelle cliniche, dati aziendali riservati, credenziali di sicurezza e documenti legali. Gli LLM, soprattutto quando incorporati nelle applicazioni, rischiano di esporre dati sensibili, algoritmi proprietari o dettagli riservati tramite il loro output. Ciò può comportare accesso non autorizzato ai dati, violazioni della privacy e violazioni della proprietà intellettuale.
Supply Chain	Le catene di fornitura LLM sono soggette a varie vulnerabilità, che possono influire sull'integrità dei dati di formazione, dei modelli e delle piattaforme di distribuzione. Esse possono causare output distorti, violazioni della sicurezza o guasti del sistema. Mentre le vulnerabilità software tradizionali si concentrano su problemi come difetti del codice e dipendenze, quì i rischi si estendono anche a modelli e dati pre-addestrati da terze parti. Questi elementi esterni possono essere manipolati tramite attacchi di manomissione (tampering) o di avvelenamento (poisoning).
Data and Model Poisoning (Avvelenamento dei dati e del modello)	L'avvelenamento dei dati si verifica quando i dati di pre-addestramento, messa a punto o incorporamento vengono manipolati per introdurre vulnerabilità, backdoor o pregiudizi. Questa manipolazione può compromettere la sicurezza, le prestazioni o il comportamento etico del modello, portando a output dannosi o a capacità compromesse. L'avvelenamento dei dati può colpire diverse fasi del ciclo di vita LLM, tra cui la pre-formazione (apprendimento da dati generali), la messa a punto (adattamento dei modelli a compiti specifici) e l'incorporamento (conversione del testo in vettori numerici). L'avvelenamento dei dati è considerato un attacco all'integrità poiché la manomissione dei dati di addestramento incide sulla capacità del modello di fare previsioni accurate. I rischi sono particolarmente elevati con fonti di dati esterne, che possono contenere contenuti non verificati o dannosi.
Improper Output Handling (Gestione impropria dell'output)	La gestione impropria dell'output si riferisce specificamente a una convalida, sanificazione e gestione insufficienti degli output generati da grandi modelli linguistici prima che vengano trasmessi a valle ad altri componenti e sistemi. Poiché il contenuto generato da LLM può essere controllato tramite input rapido, questo comportamento è simile alla fornitura agli utenti di un accesso indiretto a funzionalità aggiuntive. Lo sfruttamento riuscito di una vulnerabilità di gestione impropria dell'output può causare XSS (cross-site scripting, vulnerabilità dei siti web che consente ad un attaccante remoto di "iniettare" script dannosi causa insufficienti/improri controlli dell'input nei form) e CSRF (Cross-Site Request Forgery, vulnerabilità a cui sono esposti i siti web dinamici quando sono progettati per ricevere richieste da un client senza meccanismi per controllare se la richiesta sia stata inviata intenzionalmente oppure no. Diversamente da XSS, che sfrutta la fiducia di un utente in un particolare sito, il CSRF sfrutta la fiducia di un sito nel browser di un utente) nei browser web, nonché SSRF (Server-side request forgery, vulnerabilità che consente a un aggressore di far sì che l'applicazione lato server effettui richieste a servizi/applicazioni cui non potrebbe collegarsi ed accedere), escalation dei privilegi o esecuzione di codice remoto sui sistemi backend.
Excessive Agency (Agenzia eccessiva)	Vulnerabilità che consente di eseguire azioni dannose in risposta a output inaspettati, ambigui o manipolati da un LLM, indipendentemente da ciò che sta causando il malfunzionamento dell'LLM. Un sistema LLM dispone spesso di capacità per chiamare funzioni o interfacciarsi con altri sistemi tramite estensioni (chiamate anche strumenti, competenze o plugin da diversi fornitori) per intraprendere azioni in risposta a un prompt.
System Prompt Leakage (Perdita di prompt di sistema)	Questa vulnerabilità negli LLM si riferisce al rischio che i prompt di sistema o le istruzioni utilizzate per guidare il comportamento del modello possano contenere anche informazioni sensibili che non erano destinate a essere scoperte. I prompt di sistema sono progettati per guidare l'output del modello in base ai requisiti dell'applicazione, ma possono contenere inavvertitamente segreti. Quando vengono scoperti, queste informazioni possono essere utilizzate per effettuare attacchi. È importante comprendere che il prompt di sistema non deve essere considerato un segreto, né deve essere utilizzato come controllo di sicurezza. Di conseguenza, dati sensibili come credenziali, stringhe di connessione, ecc. non devono essere contenuti nel linguaggio del prompt di sistema.
Vector and Embedding Weaknesses (Vulnerabilità dei vettori e degli incorporamenti)	Vulnerabilità significative nei sistemi che utilizzano Retrieval Augmented Generation (RAG), tecnica di adattamento del modello che migliora le prestazioni e la pertinenza contestuale delle risposte dalle applicazioni LLM, combinando modelli linguistici pre-addestrati con fonti di conoscenza esterne. Le modoalità in cui i vettori e gli incorporamenti vengono generati, archiviati o recuperati possono essere sfruttate da azioni (intenzionali o non intenzionali) per iniettare contenuti dannosi, manipolare gli output del modello o accedere (intenzionali o non intenzionali) per iniettare contenuti dannosi, manipolare gli output del modello o accedere a informazioni sensibili.
Misinformation (Disinformazione dovuta ad errori e non intenzionale, come invece è la "disinformation")	Vulnerabilità che si verifica quando gli LLM producono informazioni false o fuorvianti che sembrano credibili. Una delle principali cause è l'allucinazione, quando l'LLM genera contenuti che sembrano accurati ma sono inventati. Le allucinazioni si verificano quando gli LLM colmano le lacune nei loro dati di formazione utilizzando modelli statistici, senza comprendere veramente il contenuto. Di conseguenza, il modello può produrre risposte che sembrano corrette ma sono completamente infondate. Oltre alle allucinazioni, altre cause di disinformazione sono i pregiudizi introdotti dai dati di formazione dell'LLM e le informazioni incomplete, dovute a un controllo inadeguato.
Unbounded Consumption (Consumo illimitato)	Questa vulnerabilità si riferisce al processo in cui un Large Language Model (LLM) genera output basati su query o prompt di input. L'inferenza (una proposizione viene derivata dalle premesse) è una funzione critica degli LLM, che implica l'applicazione di modelli e conoscenze appresi per produrre risposte o previsioni pertinenti. Il consumo illimitato si verifica quando un'applicazione Large Language Model (LLM) consente agli utenti di condurre inferenze eccessive e incontrollate, comportando rischi quali negazione del servizio (DoS), perdite economiche, furto di modelli e degradazione del servizio. Le elevate richieste computazionali degli LLM, in particolare negli ambienti cloud, li rendono vulnerabili allo sfruttamento delle risorse e all'uso non autorizzato.

Fig. 2-3

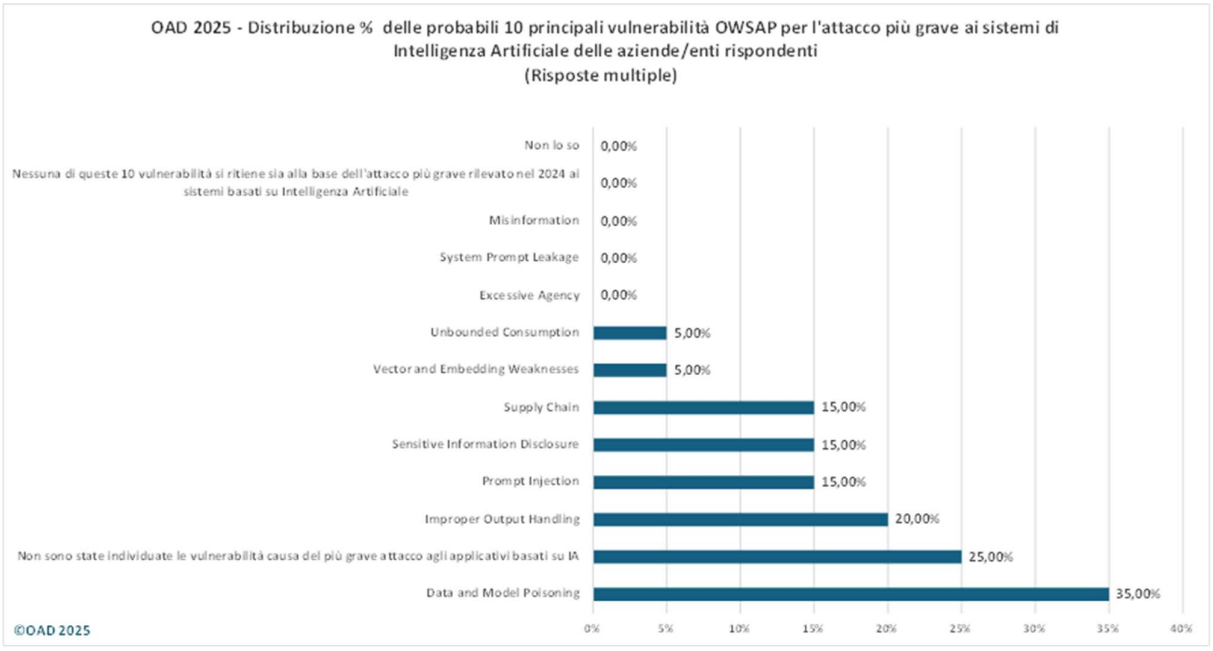


Fig. 2-4

A giudizio dell'autore, le risposte avute sono ragionevoli ed interessanti, ma sono da considerare come prime indicazioni del fenomeno degli attacchi ai sistemi IA. Da un lato sono relativamente pochi quelli che hanno rilevato attacchi ai loro sistemi di IA, dato che la maggior parte usa sistemi consolidati di tipo commerciale (ad esempio ChatGPT, Copilot, Gemini, Claude, che operano su piattaforme cloud ad alta affidabilità, disponibilità e sicurezza; dall'altra l'uso di questi sistemi è prevalentemente per ricerche e confronti di informazioni e, a parte i problemi di copyright e di privacy, i risultati sono in qualche misura facilmente verificabili).

Scenario internazionale e resilienza digitale necessaria

I risultati dell'indagine OAD 2025 di AIPSI sono sostanzialmente allineati con quanto emerge dal World Economic Forum, da ENISA, da ACN.

Il 2024 è stato caratterizzato da uno scenario internazionale fortemente instabile, con guerre ibride (Ucraina, Medio Oriente), che hanno visto un'escalation di **cyber warfare, disinformazione e attacchi digitali mirati** contro infrastrutture pubbliche e private. **Permane complessivamente l'ampia diffusione di gravi attacchi digitali e di un forte rischio cibernetico a livello mondiale, europeo e in Italia: e le prime indicazioni nel primo semestre del 2025 stanno confermando questa situazione.**

Per l'Italia,dal 2018 si è entrati nell'**era della insicurezza digitale sistemica** (systemic cyber insecurity): e questa era vale per l'intero mondo. Nonostante le misure di sicurezza implementate, queste non sono state e non sono ancora capaci di bloccare tutti gli attacchi digitali. Con le misure attualmente a disposizione, che vanno potenziate e non certo dismesse, occorre imparare a gestire questa insicurezza grazie a **politiche di resilienza dell'intera impresa e del suo sistema informativo**. Resilienza che lato business può essere garantita dalla "**business continuity**", la continuità operativa almeno dei processi fondamentali per il funzionamento dell'impresa; e lato informatico da un piano di **Disaster Recovery** effettivo, attuabile e "provato".

Note

---

[1] Liberamente scaricabile da: https://www.aipsi.org/aree-tematiche/osservatorio-attacchi-digitali/oad-2025/per-scaricare-il-rapporto-oad-2025.html

[2] AIPSI, Associazione Italiana Professionisti Sicurezza Informatica, capitolo italiano della mondiale ISSA:

[3] DESI, Digital Economy and Society Index, è un indice composito che sintetizza vari rilevanti indicatori sulle prestazioni digitali in Europa e traccia l'evoluzione dei vari membri EU nella competitività digitale.

[4] OWASP, Open Web Application Security Project, iniziativa che formula a livello mondiale linee guida, strumenti e metodologie per migliorare la sicurezza delle applicazioni prevalentemente in ambito web (https://owasp.org/)

[5] https://owasp.org/www-project-top-10-for-large-language-model-applications/

Marco R. A. Bozzetti

Presidente Onorario AIPSI – Associazione Italiana Professionisti Sicurezza Informatica