

Attacchi digitali ed Intelligenza Artificiale nell'indagine OAD 2025 di AIPSI

Marco R. A. Bozzetti

Presidente Onorario e Past President AIPSI (m.bozzetti@aipsi.org)

Ideatore e coordinatore iniziativa OAD

Founder e CEO Malabo srl (www.malboadvisoring.it)

AIPSI, capitolo italiano della mondiale ISSA

- Associazione no-profit di sole persone fisiche
- Obiettivo principale: l'indirizzamento e la crescita professionale dei suoi Soci
- Tre tipologie di Socio
 - AIPSI-ISSA (US\$ 160,00 /anno)
 - SOLO AIPSI (€ 50,00/anno)
 - AIPSI GIOVANE (fino a 26 anni: primo anno gratuito, poi € 25,00/anno)
 - Si veda:
<https://www.aipsi.org/associazione/come-associarsi.html>
- AIPSI è Socio **FIDA**Inform

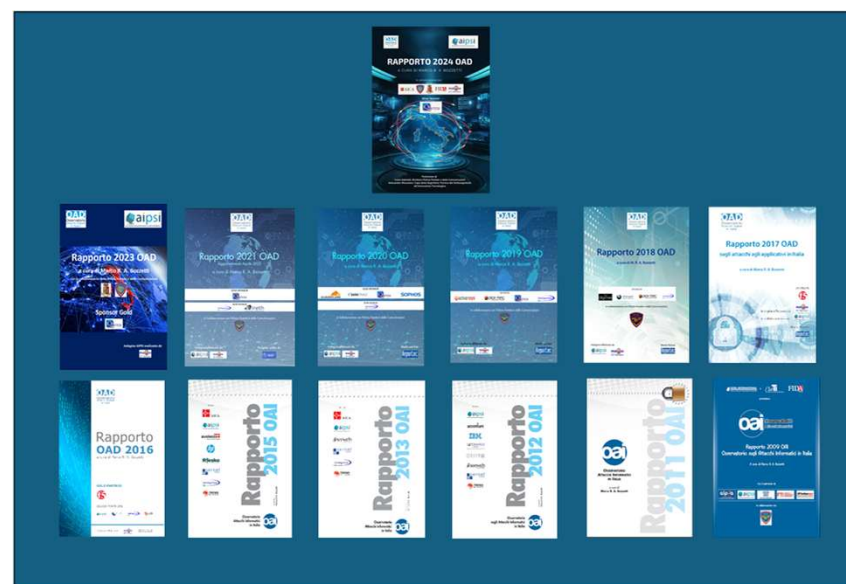


2025
20 anni AIPSI e 40 anni ISSA

- Servizi riservati a tutti i Soci AIPSI
 - **supporto alle certificazioni**, in particolare per **eCF Plus** (EN 16234-1:2016) per profili sulla sicurezza digitale con **con forti sconti**
 - **Mentorship gratuita sull'indirizzamento e sulla crescita professionale**
 - **SIG di approfondimento e discussione**
 - **Network Soci a livello nazionale**
- Servizi riservati ai Soci AIPSI-ISSA
 - **ISSA Journal**
 - **ESG ISSA Survey "The Life and Times of Cyber Security Professionals"**
 - **Convegni, workshop, webinar in inglese**
 - **Corsi online in inglese**
 - **SIG, Special Interest Group**
 - Privacy
 - Women in Security
 - **Accordi con sconti per certificazioni individuali**
 - **Network Soci a livello mondiale**

- **OAD** è l'unica indagine online in Italia (**completamente indipendente e “terza” rispetto ai vari attori in gioco**) 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 in esercizio.
- **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.

- **18 anni consecutivi di indagini**
 - basate su questionari online anonimi
 - libero accesso al questionario per tutti i referenti di un sistema informativo operante in Italia
- **Per tutti i settori merceologici + Pubbliche Amministrazioni Centrali e Locali**
- **13 Rapporti pubblicati**
- Uno **specifico sito web** che costituisce il **repository** di tutti i Rapporti OAD/OAI pubblicati e della documentazione sui vari eventi e sugli articoli pubblicati nei quali AIPSI ha presentato dati emersi dalle indagini: www.oadweb.it



OAD 2025: l'indagine diviene maggiorenne

OAD 2025
ATTACCHI DIGITALI
IN ITALIA

- Il piano di lavoro, con allegate le proposte di sponsorizzazione (in italiano e in inglese) e di patrocinio gratuito, è in <https://www.aipsi.org/aree-tematiche/osservatorio-attacchi-digitali/oad-2025.html>
- **Il questionario OAD 2025, rigorosamente anonimo, è online e compilabile:**
<https://www.aipsi.org/aree-tematiche/osservatorio-attacchi-digitali/oad-2025/questionario-oad-2025.html>

COMPILATELO (o fatelo compilare dai vostri tecnici) e PASSATE PAROLA ad aziende/enti che conoscete

La struttura del Questionario OAD 2025

S1 - Brevi informazioni sulla Azienda/Ente della/del rispondente

S2 - Attacchi digitali di ogni tipo al Sistema Informativo rilevati nell'intero 2024

S3 - Approfondimento **attacchi ai siti e alle applicazioni web** del Sistema Informativo, con riferimento alle top ten vulnerabilità individuate da OWSAP per questi ambienti

S3B - Approfondimento sugli attacchi ad **applicativi basati su Intelligenza Artificiale**, con riferimento alle top ten vulnerabilità 2025 individuate da OWSAP per questi ambienti

S3C - Approfondimento sugli **attacchi a sistemi ed apparati OT, Operation Technology**

S4 - Attacchi più temuti nel prossimo futuro

S5 - Macro caratteristiche del Sistema Informativo cui la/il rispondente fa riferimento

S6 - Misure tecniche in atto per la sicurezza digitale dell'intero Sistema Informativo

S7 - Misure organizzative di sicurezza digitale nell'Azienda/Ente della/del rispondente

S8 - Ruolo della/del rispondente

S10 – Calcoli (non visibili) e presentazione finale in tempo reale della macro valutazione del livello di sicurezza del SI

Gli elementi caratterizzanti il questionario OAD 2025

- Due sole domande sugli attacchi rilevati nel 2024 in riferimento alle tipologie di attacco ed alle famiglie di tecniche di attacco così da poter avere dati di trend generali sugli attacchi (che cosa viene attaccato e con quali tecniche) dal 2007 al 2024;
 - **Tra le tipologie di attacco la supply chain informatica**
- le **domande di approfondimento** (se nel 2024 si sono subiti attacchi in queste aree):
 - ai siti e agli ambienti web,
 - alle applicazioni basate su Intelligenza Artificiale
 - agli ambienti OT, Operational Technology, che includono anche apparati informatici medico-sanitari-chirurgici;
- Sono opzionali le domande sulle **misure di sicurezza digitale** presenti nei sistemi informativi oggetto delle risposte,
 - ma solo se si risponde a queste **domande si potrà avere, alla fine della compilazione, la macro-valutazione del livello di sicurezza digitale in essere un funzione delle risposte selezionate**
- **L'Intelligenza Artificiale è inserita anche nelle tipologie di attacco (che cosa attacco) e nelle tecniche di attacco (come attacco)**

Le (famiglie di) tipologie di attacco in OAD 2025

- Distruzione e/o compromissione FISICA di dispositivi ICT FISSI o di loro parti
- FURTO dispositivi FISSI ICT o di loro parti
- FURTO di dispositivi ICT MOBILI di proprietà dell'azienda/ente e in uso presso i suoi dipendenti/collaboratori
- FURTO INFORMAZIONI da singoli specifici sistemi FISSI ICT (PC, server, storage system, etc.) del Sistema Informativo, anche terzarizzati/in cloud
- FURTO INFORMAZIONI relative all'azienda/ente da sistemi MOBILI (palmari, smartphone, tablet, ecc.) sia di proprietà dell'azienda/ente sia dell'utente finale che li usa in logica BYOD
- Attacchi ALL'IDENTIFICAZIONE, AUTENTICAZIONE E CONTROLLO ACCESSI degli utenti finali e privilegiati
- Attacchi alle RETI locali e geografiche, fisse e wireless, inclusi i collegamenti ad Internet, e ai DNS nel corso del 2022
- Attacco e/o uso non autorizzato di SISTEMI IT NEL LORO COMPLESSO (dal PC agli host fisici e virtuali). anche terzarizzati o in cloud
- MODIFICHE malevoli e/o non autorizzate ai PROGRAMMI APPLICATIVI e alle loro configurazioni, del Sistema Informativo anche terzarizzate e in cloud
- MODIFICHE malevoli e/o non autorizzate alle INFORMAZIONI trattate dalle applicazioni del Sistema Informativo, anche quelle terzarizzate/in cloud
- SATURAZIONE (DoS, DDoS) risorse digitali del Sistema Informativo, anche quelle terzarizzate/in cloud
- Attacchi ai propri sistemi/servizi digitali in CLOUD o comunque TERZIARIZZATI presso Fornitori terzi
- Attacchi a dispositivi dei sistemi OT, OPERATIONAL TECHNOLOGY, ivi inclusi i sistemi IoT/IIoT, i sistemi per l'automazione industriale ((SCADA, DCS, PLC, ..) e la robotica
- Attacchi alla "SUPPLY CHAIN" causati da vulnerabilità di fornitori e/o clienti interconnessi
- Attacchi a sistemi/servizi/applicativi basati su Intelligenza Artificiale.

Le (famiglie di) tecniche di attacco considerate in OAD 2025

- Attacco fisico
- Raccolta malevola e non autorizzata di informazioni
- Script e programmi maligni
- Agenti autonomi
- Toolkit
- Botnet e simili
- Utilizzo di strumenti e tecniche di Intelligenza Artificiale
- Utilizzo di due o più tecniche di attacco, inclusi gli APT, Advanced Persistent Threat

IA Generativa e LLM

IA Generativa

crea contenuti come immagini, video, musica, audio e testo basandosi su **modelli di deep learning** addestrati su grandi set di dati.

- Il deep learning è una tecnica di **machine learning** per l'analisi e l'interpretazione di grandi volumi di dati basata su **reti neurali**
- Il **trasformatore** è un tipo di rete neurale in grado di elaborare il linguaggio più rapidamente, ed è usato per LLM
- I principali scenari di utilizzo dell'IA generativa sono chatbot, creazione ed editing di immagini, assistenza alla stesura di codice software, supporto alle decisioni e alla ricerca scientifica.

LLM, Large Language Model

modello di intelligenza artificiale (IA) che comprende e genera testo in grado di:

- Riconoscere e interpretare il linguaggio umano
- Tradurre, riassumere, e generare testo
- Rispondere a domande
- Analizzare il sentiment
- Riconoscere la voce
- Gli LLM sono formati analizzando grandi quantità di dati testuali e utilizzano reti neurali profonde.
- Sono usati in molti settori, come il digital marketing, la finanza, le assicurazioni, la sanità, e le risorse umane.

Esempi di servizi ed applicazioni informatiche basate su Intelligenza Artificiale che già usiamo

- ChatGPT
- Chatbot: programmi che simulano conversazioni umane → helpdesk, e-commerce, etc.
- Assistenti vocali: Siri di Apple, Cortana di Microsoft, Bixby
- Smart home (domotica): Amazon Alexa, Google Assistant
- Traduttori tra lingue straniere: testo-testo, voce-voce, voce-testo, testo-voce
- Riconoscimento facciale
- Strumenti di supporto alla guida in auto
- Machine learning (ML) nel settore finanziario e sanitario
- Creazione di immagini
- IA a supporto/alla base di innovativi strumenti di sicurezza digitale

Esempi uso IA per attacchi digitali

- Phishing, spear phishing, deepfake phishing
- Deepfake
 - Clonazione del video e della voce
- Creazione false identità digitali e falsi documenti
- Uso di IA per social engineering
- Uso di IA per analisi vulnerabilità
- Uso di IA per modifiche (continue e in tempo reale) al codice di malware per non farlo identificare
-

Top Ten vulnerabilità in applicazioni IA 2025 di OWASP (LMM e Generative AI) - 1

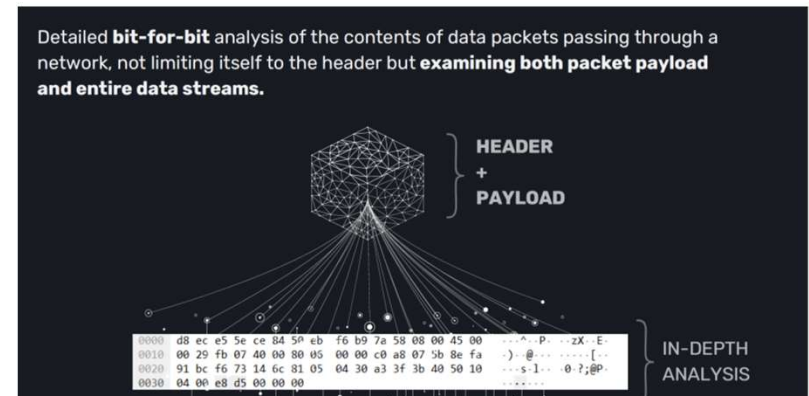
- **Prompt Injection (*iniezione di prompt*)**: questa vulnerabilità si verifica quando i prompt dell'utente alterano il comportamento o l'output dell'LLM in modi non intenzionali, causando potenzialmente la violazione delle linee guida, la generazione di contenuti dannosi, l'abilitazione di accessi non autorizzati o l'influenza di decisioni critiche.
- **Sensitive Information Disclosure (*Divulgazione di informazioni sensibili*)**, che dovrebbero essere protette in accordo all'attuale normativa europea sulla privacy, il GDPR. 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. Mentre le vulnerabilità software tradizionali si concentrano su problemi come difetti del codice e dipendenze, qui 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. 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.
- **Improper Output Handling (*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/impropri controlli dell'input nei form) e CSRF (Cross-Site Request Forgery, vulnerabilità a cui sono esposti i siti web dinamici).

Top Ten vulnerabilità in applicazioni IA 2025 di OWASP (LLM e Generative AI) - 2

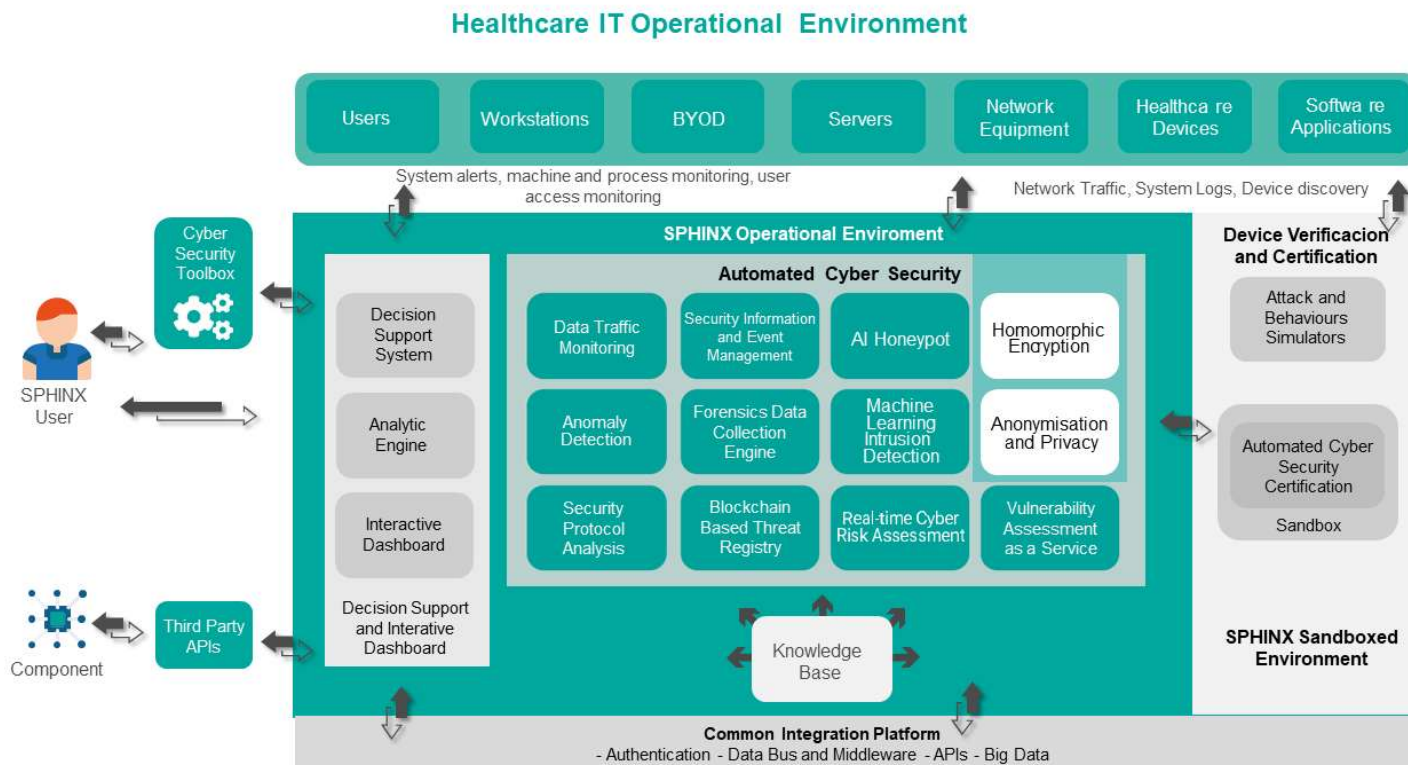
- **Excessive Agency (*agenzia eccessiva*)**: vulnerabilità che consente di eseguire azioni dannose in risposta a output inaspettati, ambigui o manipolati da un LLM, indipendentemente da ciò che sta causando il malfunzionamento dell'LLM. Un sistema LLM dispone spesso di capacità per chiamare funzioni 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. Quando vengono scoperti, queste informazioni possono essere utilizzate per effettuare attacchi.
- **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.
- **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.
- **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. Il consumo illimitato si verifica quando un'applicazione 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.

LECS: un esempio tutto italiano di una famiglia di appliance di sicurezza digitale basata su 3 motori di IA

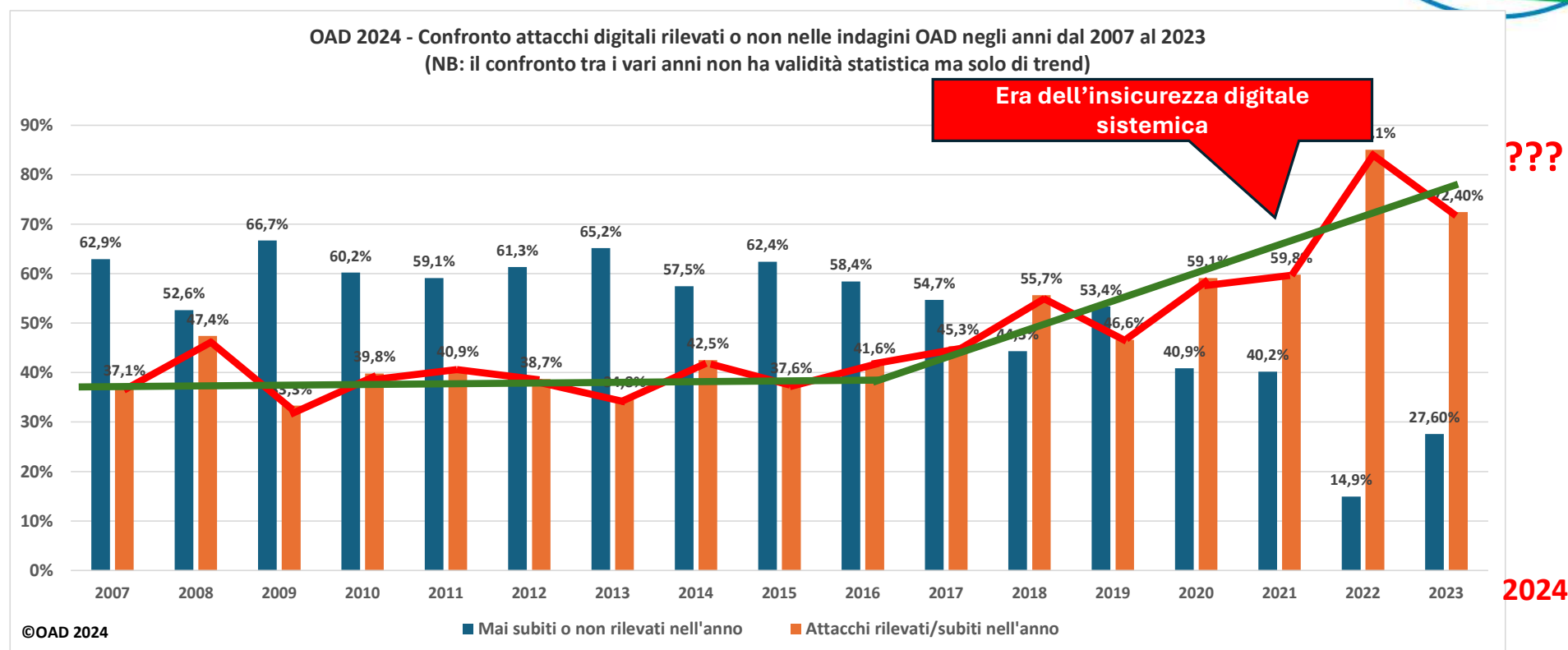
- Appliance *plug&play* con funzionalità di **NDR** (Network Detection and Response) e di **IPS** (Intrusion Prevention System) che protegge qualsiasi rete LAN, infrastrutture ICT e impianti industriali (OT)
 - Deep inspection di tutto il traffico (L3 e L2) nella LAN e sulle connessioni ad Internet
- Si avvale di una concatenazione di **3 differenti motori IA** per analizzare ogni pacchetto in transito:
 - **SPECTO** Detecton Engine
 - **TIRESIA** Threat Forecast
 - **RAISES** Autonomous Response
- Analizza e registra il comportamento dei vari dispositivi collegati (ML) ed autonomamente interviene in caso di grave minaccia creando pacchetti che si mescolano a quelli della minaccia, neutralizzandola



Progetto UE per un cybersecurity toolkit basato su IA per il settore sanitario



Il trend degli attacchi digitali dalle indagini OAD 2007-2023



GRAZIE PER L'ATTENZIONE e ...

Questa presentazione sarà scaricabile in pdf dal sito web di AIPSI

**COMPILATE e FATE COMPILARE il QUESTIONARIO OAD 2025 e
PASSATE PAROLA ai vostri interlocutori:**

<https://www.aipsi.org/aree-tematiche/osservatorio-attacchi-digitali/oad-2025/questionario-oad-2025.html>



<https://www.aipsi.org/>

Se volete essere sistematicamente aggiornati e **crescere professionalmente** nel campo della sicurezza digitale:

- **Diventate Soci di AIPSI**
 - <https://www.aipsi.org/associazione/come-associarsi.html>



<https://www.issa.org/>