MODULO SICUREZZA INFORMATICA

Big Data e Machine Learning per la Cyber Threat Intelligence

Metodologie, architetture e sfide aperte

Vincenzo Calabrò

IL PROBLEMA

L'offesa si è fatta industria

I numeri che aprono il corso descrivono un'asimmetria: l'attacco scala, la difesa manuale no.

> 450.000

nuovi malware e applicazioni indesiderate ogni giorno (AV-TEST)

~ 60%

delle intrusioni iniziali parte dal phishing (ENISA ETL 2025)

> 80%

del social engineering osservato è ormai AI-supported

> 42.000

nuove vulnerabilità in un anno, weaponizzate in pochi giorni

La Cyber Threat Intelligence promette di fornire il contesto che manca agli analisti. Ma renderla praticabile oggi richiede automazione — e l'automazione ha condizioni precise.

Obiettivi di apprendimento

1 Inquadrare

Definire la CTI, il suo ciclo e i quattro livelli; collocare i framework ATT&CK, Kill Chain e Diamond Model.

2 Architettare

Descrivere la pipeline Big Data che sostiene la CTI, dalle fonti allo storage, e le sue dimensioni critiche.

3 Analizzare

Valutare i metodi ML per la detection e le tecniche NLP per l'estrazione da fonti aperte, con i loro limiti.

4 Condividere

Comprendere standard, piattaforme e vincoli normativi dello scambio di intelligence (STIX/TAXII, NIS2).

5 Attribuire

Ragionare sull'attribuzione come processo a confidenza graduata e sui suoi limiti epistemici.

6 Difendere il modello

Riconoscere le vulnerabilità adversarial dei sistemi ML e le direttrici di ricerca aperte.

PARTE I

Fondamenti di Cyber Threat Intelligence

Ciclo, livelli, attori e framework di analisi strutturata

Il ciclo di intelligence

1 Dato → Informazione → Intelligence

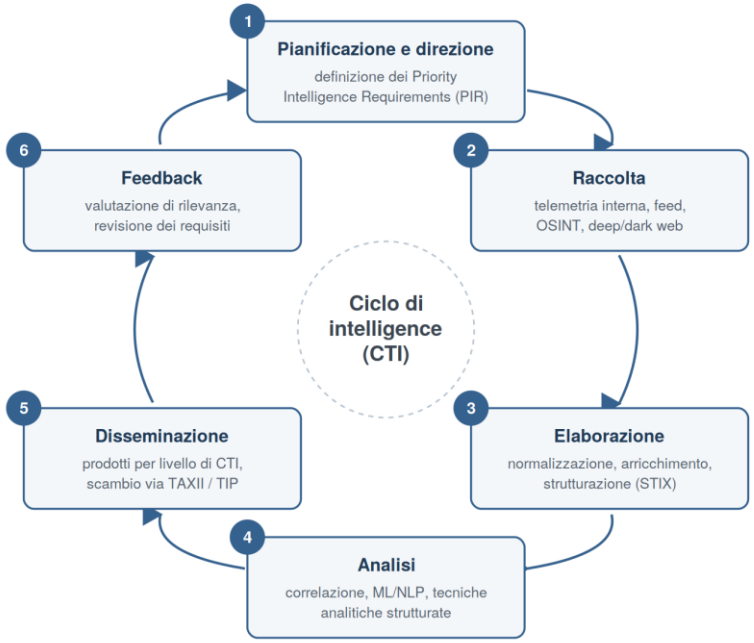
La progressione che definisce la disciplina: dall'osservabile grezzo al prodotto su cui si decide.

2 Sei fasi

Direzione, raccolta, elaborazione, analisi, disseminazione, feedback. L'analisi è dove il ML incide di più.

3 Un modello idealizzato

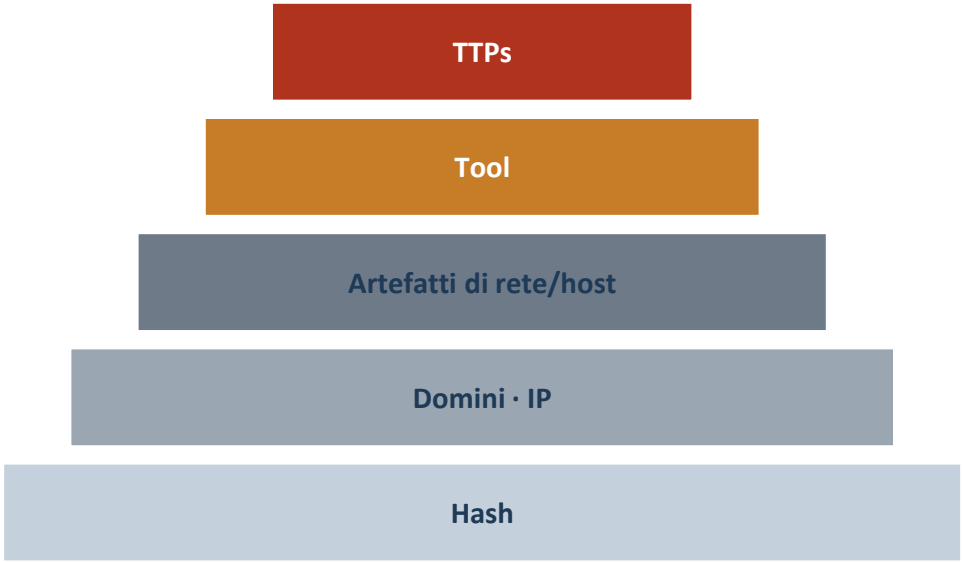
Hulnick (2006): nella pratica le fasi procedono in parallelo. Nel cyber l'anello si comprime a velocità di macchina.



Quattro livelli, un principio economico

Livello	Destinatari	Orizzonte	Contenuto
Strategica	vertici, CISO	mesi–anni	trend, attori, geopolitica
Tattica	security architect	settimane	TTPs, mappatura ATT&CK
Operativa	SOC, IR, hunter	giorni	campagne, infrastrutture
Tecnica	strumenti automatici	ore–giorni	IoC: hash, IP, domini

Pyramid of Pain (Bianco, 2013)



↑ *costo per l'avversario* · ↑ *difficoltà di estrarla automaticamente*

Principio chiave. Il valore dell'intelligence cresce con il livello di astrazione, ma cresce in pari misura il costo di produrla automaticamente: estrarre un hash è banale, estrarre TTPs da un report richiede NLP (Parte III).

Attori della minaccia e TTPs

TTP: l'unità dell'intelligence tattica

Tactics — il perché, l'obiettivo tattico
Techniques — il come, i modi per conseguirlo
Procedures — le implementazioni osservate

Specificità crescente, volatilità crescente: le procedure cambiano a ogni campagna, le tattiche restano per anni.

State-nexus (APT)	spionaggio, sabotaggio; risorse elevate
Criminalità organizzata	profitto; filiera RaaS, initial access broker
Hacktivism	ideologia; DDoS e defacement ad alta visibilità
Insider / opportunisti	capacità bassa, motivazioni eterogenee

Attenzione: la tassonomia è una lente, non una gabbia

Volume ≠ impatto. Nell'ETL 2025 il 79,4% degli incidenti è ideologico (DDoS hacktivisti), ma solo il 2% causa vera interruzione; il ransomware, pur in calo dell'11%, resta la minaccia a più alto impatto.

Convergenza. Gruppi hacktivisti adottano ransomware (es. FunkSec); il riuso di strumenti tra insiemi di intrusione diversi diluisce i confini. Chi addestra modelli su dati di incidenti deve saper trattare l'ibridazione.

Tre framework, tre prospettive complementari

Cyber Kill Chain

Lockheed Martin, 2011

Prospettiva temporale-sequenziale: sette fasi dell'intrusione. Difesa intelligence-driven — basta spezzare la catena in un punto.

Limite: Linearità; impronta perimetrale e malware-centrica.

Diamond Model

Caltagirone et al., 2013

Prospettiva relazionale: l'evento come relazione fra adversary, capability, infrastructure, victim. Abilita il pivoting analitico.

Limite: Poco prescrittivo sul piano operativo.

MITRE ATT&CK

MITRE, dal 2013

Knowledge base empirica di comportamenti avversari. Lingua franca de facto: spazio di etichette per detection e NLP.

Limite: Copertura sbilanciata sull'osservato; non sequenziale.

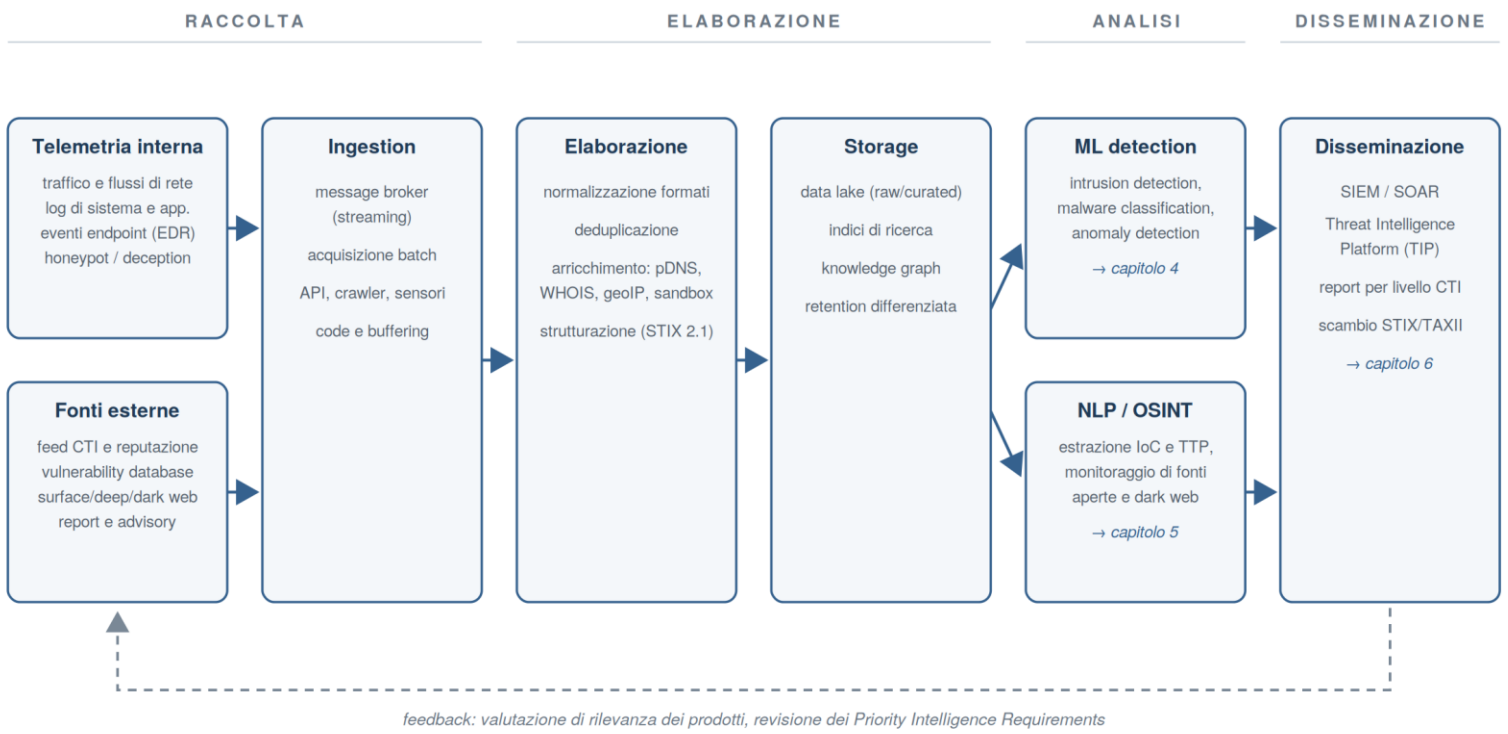
Non alternativi ma componibili: la Kill Chain dà l'asse temporale, il Diamond la struttura relazionale, ATT&CK il vocabolario comportamentale.

PARTE II

Big Data per la CTI

Fonti, architetture e la dimensione che nessun altro dominio possiede

La pipeline di riferimento



Le dimensioni del Big Data

Volume
TB/giorno di telemetria + 450k campioni/die

Velocità
streaming continuo; l'intelligence tecnica decade in ore

Varietà
strutturato, semi-, non strutturato, binario

Veridicità
parte del dato è prodotta dall'avversario

Le macro-fasi ricalcano il ciclo di intelligence. Il feedback distingue una pipeline di intelligence da una catena di ETL.

Le fonti e il problema della qualità

Telemetria
rete, endpoint, log, cloud

Deception
honeypot: ogni interazione è sospetta

Feed CTI
IoC pronti, qualità variabile

Vuln. intel
CVE, sfruttamento attivo

Malware intel
campioni e sandbox

OSINT
report, forum, dark web (Parte III)

«Reading the Tea Leaves» (USENIX Security 2019)

Feed che dichiarano di coprire la stessa categoria di minaccia mostrano sovrapposizioni sorprendentemente basse: nessuno osserva più di una frazione del fenomeno. Latenza e accuratezza confermate come problematiche da valutazioni indipendenti.

Conseguenza operativa. La qualità va gestita come proprietà di primo livello — scoring, invecchiamento degli indicatori, provenienza — perché nessun modello a valle compensa un ingresso scadente.

PARTE III

Analisi: Machine Learning e NLP

Detection sul dato strutturato, estrazione dal linguaggio naturale

Detection: paradigmi e un avvertimento storico

Supervised
richiede etichette costose, rumorose, deperibili

Unsupervised
anomaly detection: ma anomalo ≠ malevolo

Semi/Self-supervised
il compromesso pragmatico prevalente

Deep learning
representation learning, ma opaco e fragile

Sommer & Paxson (IEEE S&P 2010)

Test of Time Award 2020

Perché l'anomaly detection basata su ML è così rara nelle reti operative? Perché trovare attacchi è un compito diverso da quelli in cui il ML eccelle: il paradigma riconosce ciò che somiglia al già visto, non la novità ostile.

Costo altissimo di ogni errore · semantic gap · variabilità del traffico benigno · valutazione difficile.

Malware e concept drift. Il bersaglio è mobile: la distribuzione dei campioni cambia di continuo. TESSERACT (USENIX 2019) mostra che ignorare il drift — con partizioni che di fatto vedono il futuro — gonfia sistematicamente le prestazioni dichiarate.

Le illusioni di valutazione

Tre ragioni per diffidare del «99% di accuratezza» sui benchmark pubblici.

1 Benchmark difettosi

CICIDS2017, il dataset più citato della sua generazione, è stato rivisto da Engelen et al. (2021): errori lungo l'intera filiera, dalla generazione del traffico all'etichettatura. La versione corretta cambia i risultati.

Il progresso misurato può essere artefatto del dataset.

2 Il tempo ignorato

Partizioni sperimentali che di fatto vedono il futuro (TESSERACT). Rimuovere il bias temporale ridimensiona drasticamente le prestazioni dichiarate dai classificatori.

Leggere i risultati con l'anno del dato, non solo del metodo.

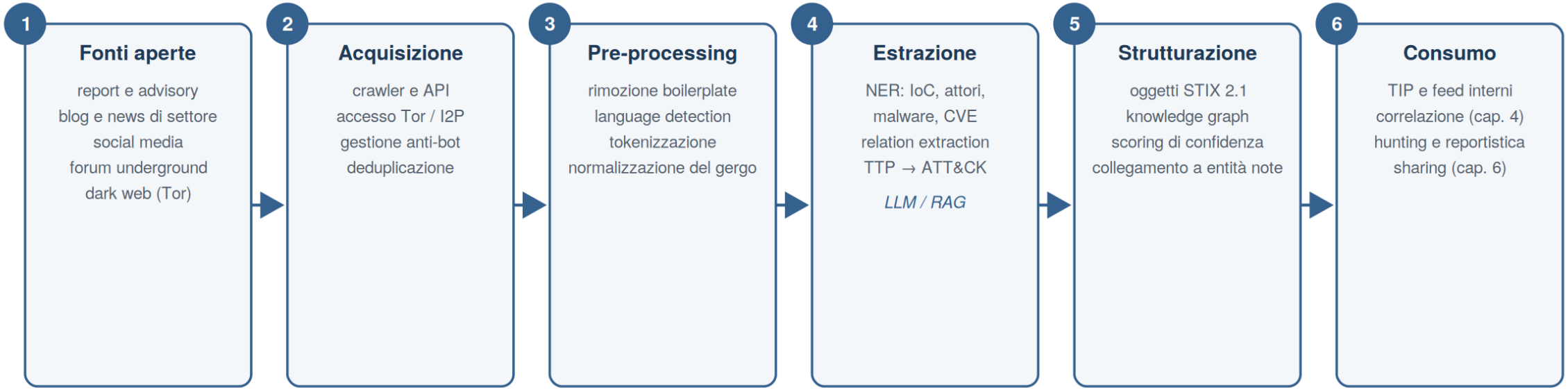
3 La base-rate fallacy

Axelsson (2000): quando gli attacchi sono rari, il valore di un allarme è dominato dai falsi positivi. Serve un FPR $\approx 10^{-5}$, spesso irraggiungibile.

La ROC-AUC resta > 0,99 mentre la precision crolla.

Arp et al. (USENIX Security 2022): dieci insidie ricorrenti, trovate pervasivamente in 30 pubblicazioni di primo livello. Il detector alza il rapporto segnale-rumore del triage; non emette verdetti.

NLP: estrarre TTPs dal linguaggio naturale



IoC atomici

iACE (CCS 2016): 900k indicatori da 71k articoli, con il loro contesto.

Entità e relazioni

NER di dominio → triple per i knowledge graph.

TTPs → ATT&CK

TTPDrill (ACSAC 2017): precision e recall > 82% su report reali.

Dark web e Large Language Model

La voce diretta dell'avversario

- Mercati di exploit e di accessi (initial access broker)
- Credenziali compromesse in circolazione
- Leak site delle operazioni ransomware
- Uso predittivo: DarkEmbed, Sabottke et al. (1,1 mld di tweet, 287k CVE)

Cautele. Accesso via Tor, anti-crawling (CAPTCHA aggirati con GAN), confini giuridici, rappresentatività, e la solita veridicità: l'avversario sa di essere osservato.

LLM: promessa e cautela

Promettono di comprimere l'intera colonna NLP: estrazione zero-shot, sintesi, KG assistiti, ancorati via RAG per ridurre le hallucination.

- Un IoC inventato è disinformazione iniettata nella difesa, non rumore
- Riservatezza: report non pubblici vs API di terze parti
- Prompt injection indiretta (Greshake et al. 2023): minaccia nativa per chi legge contenuto ostile

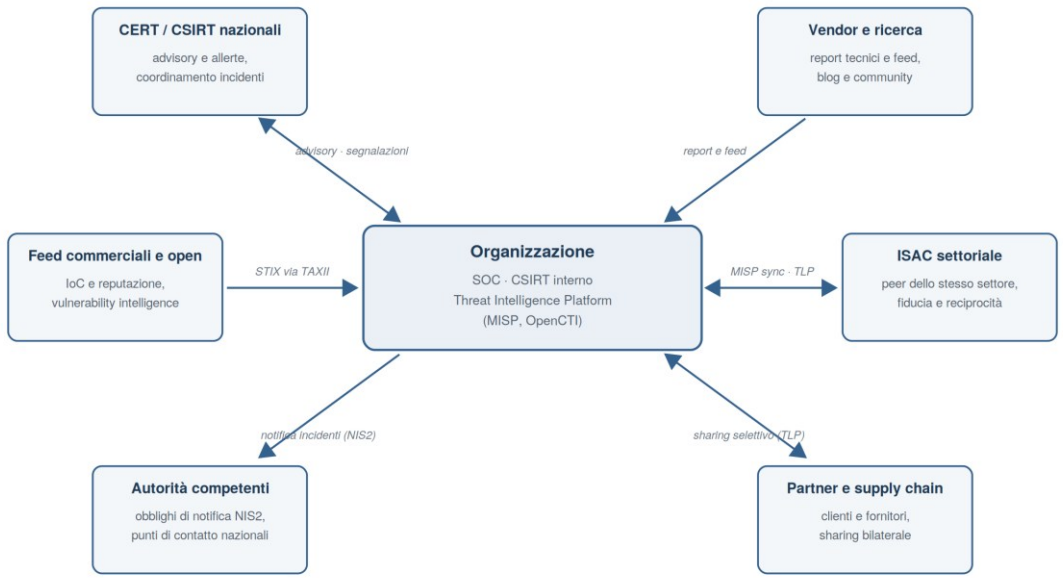
Ruolo ragionevole: copilota dell'analista, con output verificati — non estrattore autonomo in produzione.

PARTE IV–VI

Condivisione, attribuzione, robustezza

Standard e barriere, il «chi», e il modello come bersaglio

Condividere conviene. Si condivide poco, e male



STIX/TAXII · MISP/OpenCTI · ISAC, CERT/CSIRT, SOC · NIS2 (24h / 72h / 1 mese)

CTI-Lense (NDSS 2024)

≈ 6 milioni di oggetti STIX da 10 fonti pubbliche, quasi un decennio.

> 90% del materiale è firme di malware e URL

0,09% degli indicator porta regole di detection utilizzabili

19% dei dati sui threat actor è errato

37,9% di ridondanza perfino dentro i singoli provider

In termini di Pyramid of Pain: circola soprattutto ciò che vale meno. Le barriere sono in gran parte non tecniche — fiducia, timori reputazionali, privacy (TLP), free riding.

Il «chi»: un processo a confidenza graduata

Sponsor / mandante

per conto di chi, con quale intento

evidenza raramente solo tecnica

Operatore

quale gruppo ha agito (intrusion set)

confidenza media, su cluster

Macchina / infrastruttura

da quali sistemi proviene

alta sul cosa, minima sul chi

Rid & Buchanan (2015)

L'attribuzione «è ciò che gli Stati ne fanno»: un processo a tre piani — tattico, operativo, strategico — con esiti in gradi di confidenza, non un verdetto.

Evidenze da pesare per contraffabilità

Code similarity e stylometry, riuso di infrastruttura, vittimologia, TTPs. Skopik & Pahi: l'affidabilità è inversa alla facilità di contraffazione.

False flag. Olympic Destroyer (2018), costruito per somigliare a più attori insieme. Il ML produce ipotesi a confidenza, mai verdetti; l'aggiudicazione resta umana.

Lo strato di ML è una superficie d'attacco

Evasion · Poisoning (incl. backdoor) · Model extraction · Inference — su assi white/black-box e targeted/untargeted.

Il vincolo del problem space

A differenza della computer vision, la perturbazione deve preservare la funzionalità: un malware evaso deve restare un eseguibile funzionante. Gli attacchi realistici costano più di quanto la letteratura tema.

«Real Attackers Don't Compute Gradients»

Apruzzese et al. (IEEE SaTML 2023): l'avversario vero preferisce le tattiche semplici — avvelenare l'ingresso incluso. Difendere l'integrità del dato conta più che blindare il modello.

SINTESI

Quattro tesi da portare a casa

1 Il valore cresce con l'astrazione

Gli IoC si aggirano in ore, le TTPs costano molto di più — ma lo sharing scambia proprio ciò che vale meno (0,09% con regole).

2 Il collo di bottiglia non è algoritmico

Feed che non si sovrappongono, benchmark difettosi, fiducia come vincolo: limiti di dato e organizzazione, non di modello.

3 Le illusioni di valutazione sono sistemiche

Base-rate fallacy, drift, dataset bias: un intero campo ha misurato per anni artefatti dei propri dataset.

4 Il modello è una superficie d'attacco

Evasion, poisoning, prompt injection — ma l'avversario reale preferisce i mezzi semplici. Contano dato e integrità.

IN SINTESI

L'automazione non sostituisce il giudizio dell'analista: ne alza il rapporto segnale-rumore.

Perché l'output di una macchina diventi intelligence — conoscenza su cui decidere — servono spiegabilità, provenienza tracciata e confidenze dichiarate. È la condizione che attraversa detection, NLP, sharing, attribuzione e robustezza.

Spunti per l'approfondimento

- Riprodurre TESSERACT su un dataset a scelta: quanto cala la performance con split temporale?
- Confrontare CICIDS2017 originale e versione corretta di Engelen et al.
- Valutare un LLM come estrattore di TTP: precisione vs rischio di hallucination.