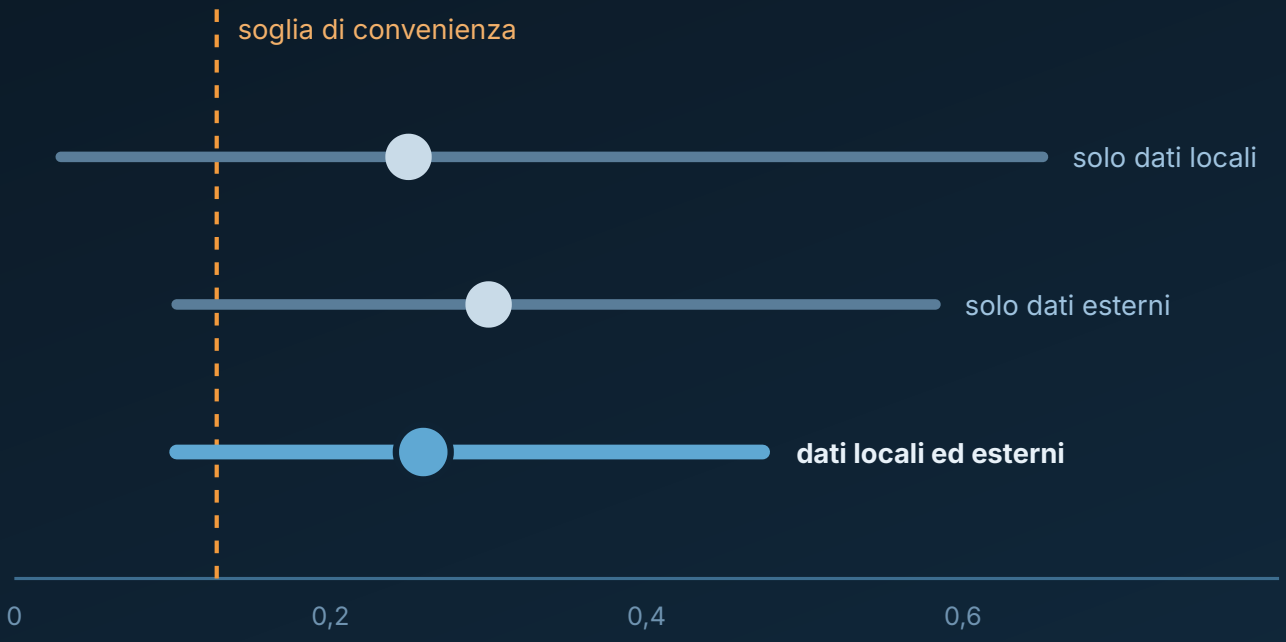
WHITE PAPER · GESTIONE DEL RISCHIO CYBER



Analisi quantitativa con pochi dati

Il contributo dell'AI-driven Cyber Risk Management

Vincenzo Calabrò

NOVEMBRE 2026

vincenzocalabro.it

WHITE PAPER · NOVEMBRE 2026

Analisi quantitativa con pochi dati

Il contributo dell'AI-driven Cyber Risk Management

Vincenzo Calabrò · 22 novembre 2026

Abstract

L'analisi quantitativa del rischio cyber richiede stime di frequenza e di perdita che una singola organizzazione riesce di rado a ricavare dai propri dati: gli eventi osservati sono pochi, una parte non viene resa pubblica e le basi informative aperte sono incomplete. In questo articolo esamino in che misura le tecniche di AI riunite sotto l'etichetta di AI-driven Cyber Risk Management possano supplire a questa scarsità. La rassegna, che ho condotto su atti normativi, pubblicazioni di organismi tecnici, report di settore e letteratura peer-reviewed disponibili a novembre 2026, individua quattro modalità di intervento: la stima di grandezze non osservabili in sede a partire da dati aggregati, il completamento di record parziali, l'estrazione di dati strutturati da testi e la combinazione formale di fonti esterne, osservazioni locali e giudizio esperto. Le prove più solide riguardano le prime due. In un esempio numerico con un solo evento osservato in sei anni, l'integrazione di dati esterni riduce del 40% l'ampiezza dell'intervallo di credibilità sulla frequenza e porta dal 67% a quasi il 90% la probabilità che un investimento di sicurezza risulti conveniente. La verifica che ho svolto su dati reali e pubblici conferma i meccanismi e ridimensiona l'esempio. Con 50 osservazioni la combinazione dimezza l'errore di stima, ma una stima esterna che sottovaluta il valore reale di almeno tre volte produce intervalli al 90% che lo contengono nel 43% dei casi, e con il contenuto informativo appreso dai dati la probabilità di convenienza scende al 52%. Un modello che completa i punteggi di severità mancanti assegna la classe corretta nel 71,5% dei casi nel primo anno e nel 54,8% nel quarto. Il contributo ha un confine, che dipende dalla ragione per cui il dato manca: quando la mancanza è legata al valore stesso del dato, come accade per le perdite non dichiarate e per gli eventi estremi, nessuna tecnica ricostruisce l'informazione assente.

Keywords: cyber risk quantification; AI-driven Cyber Risk Management; missing data; Open FAIR; EPSS; Bayesian updating; CVSS; empirical validation; NIS2.

Come citare: Calabrò, V. (2026). *Analisi quantitativa con pochi dati: il contributo dell'AI-driven Cyber Risk Management*. White paper. www.vincenzocalabro.it/article/2026/analisi-quantitativa-pochi-dati-ai-driven-cyber-risk-management.html
Testo distribuito con licenza Creative Commons Attribuzione 4.0 Internazionale (CC BY 4.0).

Indice

01	Introduzione	3
02	Fonti e metodo	4
03	L’analisi quantitativa e i dati che richiede	6
04	Dove mancano i dati	8
05	Come l’AI supplisce ai dati mancanti o scarsi	10
06	Evidenze disponibili	12
07	Un esempio: stimare il rischio con un solo evento osservato	15
08	Una verifica su dati pubblici	18
09	Vincoli normativi e standard	25
10	Ciò che l’AI non ricostruisce	27
11	Conclusioni	29
	Riferimenti	30
	L’autore	33

01 Introduzione

La direttiva NIS2 chiede che le misure di sicurezza siano proporzionate al rischio e che la proporzionalità sia valutata tenendo conto, tra l'altro, della probabilità che gli incidenti si verifichino e della loro gravità, comprese le ricadute sociali ed economiche [1]. Il regolamento DORA impone alle entità finanziarie un framework documentato di ICT risk management, che l'organo di gestione definisce, approva e sorveglia e del quale risponde [2]. Nessuna delle due norme prescrive un metodo di valutazione. Le valutazioni richieste si prestano tuttavia a essere espresse in una forma confrontabile con gli altri rischi d'impresa, cioè in termini di frequenza attesa e di entità della perdita.

Chi prova a produrre queste stime si scontra anzitutto con la scarsità dei dati. Una revisione sistematica partita da 5.219 studi ha individuato 79 dataset distinti sul rischio cyber e ha concluso che la lacuna nelle basi di dati aperte ostacola gli sforzi comuni per gestire meglio questi rischi [3]. L'estensione degli obblighi di notifica ha attenuato il fenomeno solo in parte. Nel primo semestre 2026 il Clusit ha censito nel mondo 4.911 incidenti gravi di dominio pubblico, il numero più alto mai registrato dall'associazione in un semestre; per 3.321 di essi la tecnica di attacco non si ricava dalle fonti pubbliche, contro i 943 dello stesso semestre del 2025 [4]. Sul versante delle vulnerabilità, dal 15 aprile 2026 il NIST dà la precedenza, nell'attribuzione dei punteggi di severità e degli elenchi dei prodotti interessati, alle CVE di tre categorie prioritarie e registra le altre con priorità minima, senza un'analisi programmata nell'immediato [5].

Per la singola organizzazione la difficoltà è maggiore, perché gli eventi di perdita osservati al proprio interno sono troppo pochi per ricavarne una distribuzione. I volumi da trattare crescono intanto a un ritmo che esclude la valutazione manuale. Il CVE Program ha pubblicato 48.244 record nel 2025 e 72.812 nei primi nove mesi del 2026 [6]; nel campione analizzato da Verizon lo sfruttamento di vulnerabilità è diventato il primo vettore di accesso iniziale, con il 31% dei breach [7].

Il mio obiettivo è stabilire in che misura le tecniche di AI riunite sotto l'etichetta di AI-driven Cyber Risk Management possano sostenere l'analisi quantitativa in queste condizioni. Sostengo che il loro contributo principale consista nel rendere disponibili stime là dove l'organizzazione non dispone di dati propri, secondo quattro modalità: l'uso di dati aggregati per stimare grandezze non osservabili in sede, il completamento di record parziali, l'estrazione di dati strutturati da testi e la combinazione formale di fonti eterogenee. Sostengo anche che questo contributo abbia un confine, legato alla ragione per cui il dato manca, e che trascurarlo conduca a stime più precise soltanto in apparenza. Metto poi alla prova le due tesi su dati reali e pubblici, in un dominio in cui il valore da stimare è osservabile.

02 Fonti e metodo

L'articolo combina una rassegna narrativa strutturata con una verifica su dati pubblici. La selezione delle fonti non segue un protocollo replicabile su database bibliografici, come richiederebbe una systematic review, e riflette quindi il mio giudizio; i criteri che ho adottato sono comunque dichiarati. Ho considerato atti normativi e standard nella versione vigente a novembre 2026, pubblicazioni di agenzie e organismi tecnici, report di settore con metodologia e campione dichiarati, articoli apparsi su riviste o in atti di conferenze con revisione paritaria. Ho escluso i documenti commerciali privi di metodologia e i dati riportati senza indicazione dell'origine.

Ho controllato ogni cifra sul documento che la pubblica, con le eccezioni che seguono. I valori annuali del CVE Program provengono da tabelle di terzi che citano la pagina ufficiale delle metriche, che non ho potuto consultare in forma testuale; l'ordine di grandezza trova riscontro nei dati di ENISA, che indica oltre 48.000 nuove CVE nel 2025 [8], e in quelli del NIST [5]. Ho ripreso i termini fissati dalle determinazioni dell'Agenzia per la cybersicurezza nazionale dalla sintesi di uno studio legale, indicata nei riferimenti. Ho letto gli articoli della NIS2 in una riproduzione non ufficiale del testo, il regolamento DORA e l'AI Act nelle sintesi pubblicate da EUR-Lex e, per l'art. 15 dell'AI Act, nella pagina dell'AI Act Service Desk della Commissione europea. Per la letteratura scientifica ho controllato i dati bibliografici e l'abstract, e ho tenuto le affermazioni attribuite ai singoli lavori entro ciò che l'abstract consente di sostenere. Per il benchmark CTIBench e per il rapporto sui dati sintetici ho consultato anche i passaggi del testo richiamati; del lavoro di Wang, Neil e Fenton [9] ho verificato soltanto il titolo e i dati bibliografici.

Le fonti hanno peso probatorio diverso, come riassume la Tabella 1. I report di settore dispongono di campioni ampi ma osservazionali e sono prodotti da soggetti con un interesse commerciale nel tema. La letteratura scientifica è più rigorosa nel metodo, copre porzioni ristrette del problema e arriva con alcuni anni di ritardo rispetto alla pratica.

Tabella 1. Fonti empiriche principali e relativi limiti.

Fonte	Tipo	Base dati	Limite principale
Verizon DBIR 2026 [7]	Report di settore	Oltre 31.000 incidenti, di cui oltre 22.000 breach confermati, in 145 paesi	Dati forniti da contributor volontari; campione non casuale
IBM e Ponemon, Cost of a Data Breach 2026 [10]	Report di settore	602 organizzazioni colpite tra marzo 2025 e febbraio 2026	Confronti tra gruppi non causali; sponsor commerciale
Clusit, primo semestre 2026 [4]	Rapporto associativo	4.911 incidenti gravi di dominio pubblico	Solo incidenti noti; tecnica di attacco non ricavabile dalle fonti pubbliche in due casi su tre
FIRST, modello EPSS [11]	Dato tecnico pubblico	139.473 CVE con punteggio CVSS 3.x al 1° ottobre 2023	Exploitation rilevata dai sensori dei partner
NIST, National Vulnerability Database [5]	Fonte istituzionale	Quasi 42.000 CVE arricchite nel 2025	Dal 2026 arricchimento limitato alle categorie prioritarie
Cremer et al. [3]	Revisione sistematica	5.219 studi, 79 dataset	Pubblicata nel 2022; non copre i dataset successivi

La verifica empirica usa soltanto dati pubblici, identificati dalla versione, e la descrivo insieme alle fonti nella sezione dedicata. Ho rieseguito i calcoli a partire dai file originali, con risultati identici. Metto a disposizione su richiesta il codice e i risultati numerici.

03 L'analisi quantitativa e i dati che richiede

La norma ISO/IEC 27005:2022 descrive la gestione del rischio per la sicurezza delle informazioni come un ciclo che comprende l'identificazione, l'analisi e la ponderazione del rischio, il trattamento, il monitoraggio e il riesame [12]. Il NIST Cybersecurity Framework 2.0 organizza contenuti analoghi in sei Function e colloca al centro quella di governo [13]. Nella pratica la valutazione è spesso affidata a una risk matrix, che incrocia classi ordinali di probabilità e di impatto.

Cox ha mostrato che le risk matrix hanno scarsa risoluzione, possono assegnare valutazioni qualitative più alte a rischi quantitativamente minori e, quando frequenza e gravità sono correlate negativamente, portare a decisioni peggiori di una scelta casuale [14]. Woods e Böhme, sistematizzando gli studi empirici sul danno cyber, hanno mostrato che le relazioni statistiche semplici producono risultati spuri, come una spesa in sicurezza più alta associata a tassi di compromissione maggiori, e che, quando si tiene conto di minaccia ed esposizione, gli indicatori del livello di sicurezza diventano fattori importanti nello spiegare le differenze tra i tassi di compromissione [15]. L'analisi quantitativa risponde a queste critiche esprimendo il rischio come distribuzione di perdite. Lo standard Open FAIR lo scompone in due fattori, la frequenza degli eventi di perdita e la magnitudo della perdita, ricavati a loro volta dalla frequenza degli eventi di minaccia, dalla vulnerabilità e dalle componenti primaria e secondaria del danno [16]. Wang, Neil e Fenton ne hanno proposto un'implementazione ed estensione mediante Bayesian network [9].

Con AI-driven Cyber Risk Management intendo qui l'impiego di tecniche di AI (supervised e unsupervised learning, natural language processing, large language model, modelli probabilistici alimentati da dati) a supporto di una o più fasi di questo processo. La bozza del Cyber AI Profile del NIST distingue l'uso dell'AI a fini di difesa dalla protezione dei sistemi di AI e dal contrasto degli attacchi condotti con l'AI [17]. Mi occupo del primo ambito e, al suo interno, della produzione delle stime che alimentano il modello quantitativo.

Un modello di questo tipo è esigente in fatto di dati. Per ciascun fattore occorre una stima espressa come distribuzione, e per ciascuna stima occorrono osservazioni che l'organizzazione di norma non possiede. La Figura 1 associa a ogni fattore la lacuna più frequente e il contributo che le tecniche di AI possono offrire; la Tabella 2 confronta tre modi di condurre la valutazione.

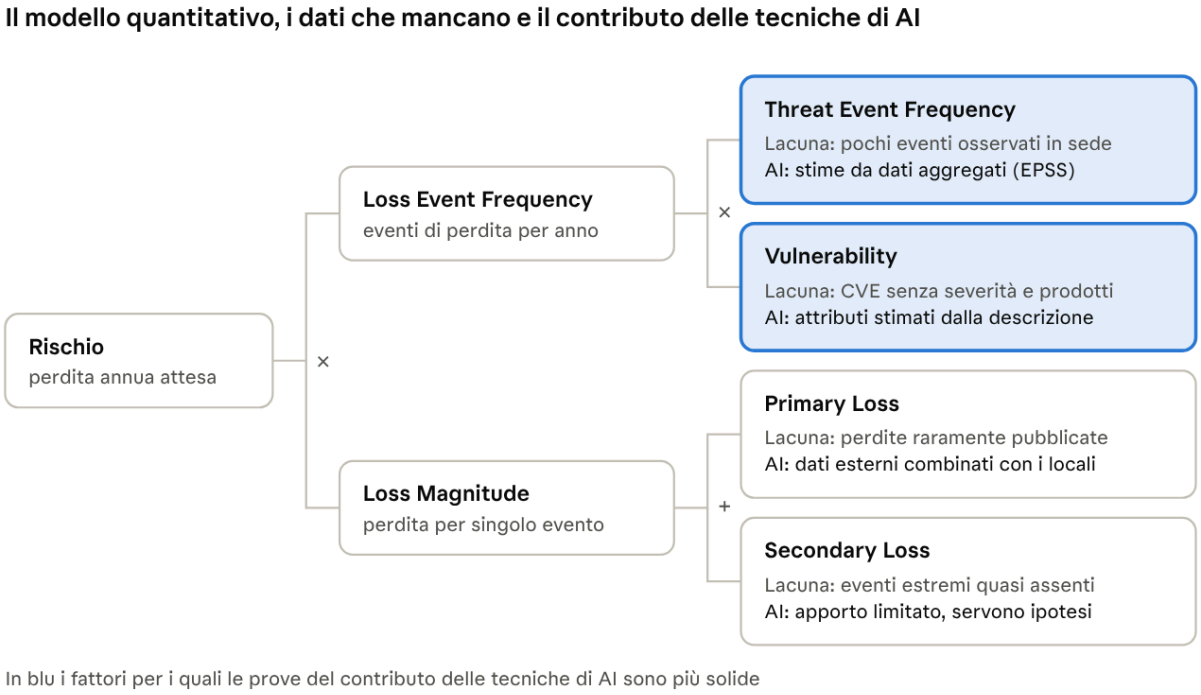


Figura 1. Fattori del modello Open FAIR, lacune di dati più frequenti e contributo delle tecniche di AI. Elaborazione dell'autore su [16].

Tabella 2. Tre modi di valutare il rischio cyber. Elaborazione dell'autore su [12], [14], [15], [16].

Aspetto	Valutazione qualitativa	Quantificazione manuale	Quantificazione AI-driven
Scala di misura	Ordinale (risk matrix)	Distribuzioni di frequenza e di perdita	Distribuzioni di frequenza e di perdita
Origine delle stime	Giudizio esperto	Giudizio esperto calibrato e pochi dati interni	Dati aggregati esterni, dati interni e giudizio esperto combinati
Trattamento dei dati mancanti	Implicito nel giudizio	Ipotesi dell'analista	Stima, completamento ed estrazione con modelli dichiarati
Aggiornamento	Annuale o semestrale	A ogni revisione dello scenario	Al variare dei dati di ingresso
Incertezza sulla stima	Non espressa	Intervalli soggettivi	Intervalli calcolati, più stretti al crescere dell'evidenza
Scenari gestibili	Molti, a bassa risoluzione	Pochi per anno	Molti, con ricalcolo automatico
Rischio metodologico	Soggettività, scarsa risoluzione	Costo elevato, dipendenza dall'analista	Qualità dei dati di ingresso, opacità dei modelli

04 Dove mancano i dati

Frequenza degli eventi

Una singola organizzazione osserva al proprio interno un numero di eventi di perdita rilevanti troppo basso per stimarne la frequenza con precisione: con campioni di poche unità l'intervallo di incertezza copre più di un ordine di grandezza, come mostra l'esempio riportato più avanti. Il ricorso a dati esterni è ostacolato dal fatto che gli incidenti non vengono resi pubblici in modo sistematico e, quando lo sono, mancano spesso delle informazioni necessarie per classificarli. Gli autori del Rapporto Clusit hanno chiamato «paradosso della trasparenza» la situazione prodotta dagli obblighi di notifica: emergono più incidenti, ma la notifica è rivolta alle autorità e non al pubblico, e le organizzazioni tendono a tenere riservati i dettagli tecnici, per ragioni di reputazione e per il rischio di subire altri attacchi [4].

Entità delle perdite

I dati sulle perdite sono ancora più rari di quelli sugli eventi. Eling e Wirfs hanno ricavato da un database di rischio operativo i casi per la loro analisi dei costi effettivi degli eventi cyber, e osservano che questi rischi sono difficili da valutare per l'assenza di dati affidabili e di analisi approfondite [18]. Le stime più citate provengono da indagini campionarie di settore. Il report IBM e Ponemon del 2026, condotto su 602 organizzazioni, indica un costo medio globale di 4,99 milioni di dollari per data breach [10]. Il dato descrive la media di un campione e non permette di ricostruire la coda della distribuzione, dalla quale dipendono gli scenari più gravi.

Vulnerabilità ed esposizione

La National Vulnerability Database ha fornito per anni i metadati su cui si basano gli strumenti di gestione delle vulnerabilità. Dall'inizio del 2024 il NIST ha accumulato un arretrato che non è riuscito a smaltire, pur avendo arricchito nel 2025 quasi 42.000 CVE, il 45% in più rispetto a qualsiasi anno precedente. Dal 15 aprile 2026 dà la precedenza nell'arricchimento alle CVE presenti nel catalogo KEV di CISA, a quelle relative a software in uso nel governo federale e al critical software definito dall'Executive Order 14028. Le altre restano pubblicate con l'indicazione di priorità minima e sono analizzate soltanto su richiesta, secondo le risorse disponibili, mentre l'arretrato con data di pubblicazione anteriore al 1° marzo 2026 è stato classificato come non programmato. L'ente ha anche annunciato che di regola non assegnerà più un proprio punteggio di severità quando la CVE Numbering Authority (CNA) che ha pubblicato la CVE ne ha già fornito uno [5]. Una quota crescente delle vulnerabilità arriva quindi agli strumenti di analisi senza l'elenco dei prodotti interessati, che l'ente in precedenza attribuiva a tutte, e con il solo punteggio di chi l'ha pubblicata.

Efficacia dei controlli

Un quarto vuoto riguarda il legame tra controlli ed esiti. Woods e Böhme mostrano che l'effetto dei controlli emerge dai dati quando si tiene conto di minaccia ed esposizione [15], informazioni di cui una singola organizzazione raramente dispone. Per l'analisi quantitativa ne consegue che l'effetto at-

teso di un investimento, cioè il parametro da cui dipende la convenienza della misura, è quasi sempre un'ipotesi.

Perché il dato manca

Le lacune descritte hanno natura diversa. Nella terminologia che discende dal lavoro di Rubin [19], un dato manca in modo completamente casuale quando la sua assenza non dipende da alcuna caratteristica dell'unità, in modo casuale quando dipende soltanto da caratteristiche osservate, in modo non casuale quando dipende dal valore che non si è potuto osservare. Una CVE priva di punteggio di severità perché estranea alle categorie prioritarie rientra nel secondo caso: l'assenza dipende da attributi noti e il valore può essere stimato dalle informazioni disponibili, a cominciare dalla descrizione testuale. Una perdita non dichiarata perché elevata o dannosa per la reputazione rientra nel terzo: l'assenza dipende dalla grandezza che si vorrebbe conoscere, e una procedura di completamento che ignori questo meccanismo tende a sottostimare le perdite. La distinzione, riassunta nella Tabella 3, delimita ciò che è lecito attendersi dalle tecniche che esamino di seguito.

Tabella 3. Fattori del modello quantitativo, dati richiesti e lacune. Elaborazione dell'autore.

Fattore (Open FAIR)	Dato richiesto	Fonte tipica	Lacuna	Meccanismo prevalente
Threat Event Frequency	Frequenza dei tentativi rilevanti	Log interni, threat intelligence	Pochi eventi per organizzazione; incidenti pubblici privi di dettagli [4]	Campione piccolo; non casuale per gli incidenti non dichiarati
Vulnerability	Probabilità che un tentativo abbia successo	Vulnerability scan, NVD	CVE prive di severità e di prodotti interessati [5]	Casuale, dipendente da attributi osservati
Primary loss	Costo diretto per evento	Contabilità interna, indagini di settore	Pochi casi interni; medie campionarie senza distribuzione [10], [18]	Presumibilmente non casuale
Secondary loss	Sanzioni, contenziosi, danno reputazionale	Casi pubblici	Eventi rari, con esiti noti dopo anni	Presumibilmente non casuale
Effetto dei controlli	Riduzione di frequenza o di magnitudo	Studi empirici	Relazioni semplici fuorvianti senza dati su minaccia ed esposizione [15]	Variabili di controllo di norma non rilevate

05 Come l'AI supplisce ai dati mancanti o scarsi

Stimare ciò che non si osserva in sede

Nessuna organizzazione dispone di osservazioni proprie sufficienti a stimare la probabilità che una determinata vulnerabilità venga sfruttata. L'Exploit Prediction Scoring System (EPSS), sviluppato in ambito FIRST, ricava questa probabilità da un modello di machine learning addestrato sull'attività di exploitation rilevata dai sensori di più partner, e pubblica ogni giorno, per ciascuna CVE, la probabilità di osservare tentativi di exploitation nei trenta giorni successivi [11], [20]. Il modello mette a disposizione di ogni utilizzatore un'informazione che esiste soltanto in forma aggregata. Lo stesso vale per il tasso di base: delle 139.473 CVE considerate da FIRST al 1° ottobre 2023, 3.852 hanno mostrato attività di exploitation nel mese seguente [11], una proporzione che una singola organizzazione non avrebbe modo di misurare.

Completare i record parziali

Il secondo intervento riguarda gli attributi assenti in record altrimenti disponibili. Shahid e Debar hanno addestrato un classificatore per ciascuna metrica del vettore CVSS a partire dalla sola descrizione testuale della vulnerabilità, e riferiscono un'accuratezza elevata e punteggi di severità molto vicini a quelli assegnati da un esperto [21]. Il benchmark CTIBench include tra i compiti proposti ai large language model la previsione del vettore CVSS e l'attribuzione della categoria CWE a partire dalla descrizione di una CVE [22]. Dopo la decisione del NIST questi metodi possono coprire un vuoto che in precedenza era colmato dagli analisti dell'ente.

La ricerca ha affrontato il problema dei campioni piccoli e incompleti anche in termini generali. TabPFN, un modello pre-addestrato su milioni di dataset sintetici, ottiene su tabelle fino a 10.000 righe risultati superiori ai metodi precedenti e accetta valori mancanti in ingresso [23]. Il risultato non riguarda dati di sicurezza; indica tuttavia che la dimensione ridotta del campione, tipica dei registri interni degli incidenti, non preclude di per sé l'uso di modelli predittivi.

Ricavare dati strutturati dai testi

Molte informazioni utili alla stima della frequenza esistono soltanto in forma narrativa, nei rapporti di threat intelligence, nei bollettini e nei resoconti di incidenti. CTIBench valuta i large language model anche sull'estrazione delle tecniche MITRE ATT&CK dai rapporti e sull'attribuzione delle campagne a un attore [22]. Applicata con sistematicità, l'estrazione trasforma un archivio di testi in una serie storica di eventi classificati, utilizzabile per stimare la frequenza degli eventi di minaccia. Gli autori del benchmark ricordano che su affidabilità, accuratezza e hallucination di questi modelli restano riserve, e ciò impone di verificare a campione i dati estratti prima di impiegarli.

Combinare fonti esterne, dati locali e giudizio esperto

La quarta modalità è la più vicina alla statistica tradizionale. Quando i dati locali sono pochi, una stima a priori ricavata da fonti esterne viene aggiornata con le osservazioni disponibili, e il peso delle due componenti dipende dalla quantità di informazione che ciascuna contiene. L'implementazione di Open FAIR mediante Bayesian network proposta da Wang, Neil e Fenton va in questa direzione [9].

Angelelli e colleghi hanno proposto un quadro statistico per stabilire la priorità delle vulnerabilità quando la conoscenza dell'insieme delle vulnerabilità esistenti è parziale, e lo hanno provato su dati simulati e su dati reali [24]. In questa modalità le tecniche di AI operano a monte, perché producono le stime esterne e gli attributi completati che entrano nel modello.

Dati sintetici

I dati sintetici sono talvolta presentati come rimedio generale alla scarsità. Il rapporto che Jordon e colleghi hanno redatto su incarico della Royal Society li considera uno strumento utile e precisa che non sostituiscono i dati reali e che ogni inferenza condotta su di essi comporta rischi aggiuntivi [25]. Nell'analisi quantitativa del rischio trovano impiego nell'addestramento e nel collaudo dei modelli, ad esempio per bilanciare classi rare. Non aggiungono informazione sulla frequenza o sull'entità delle perdite reali, poiché riproducono la distribuzione da cui sono stati generati. La Tabella 4 riassume le cinque situazioni e aggiunge quella per la quale non esiste una tecnica risolutiva.

Tabella 4. Lacune di dati e tecniche di AI: che cosa aggiungono e a quali condizioni.

Lacuna	Tecnica	Che cosa aggiunge	Evidenza	Condizione di validità
Probabilità di exploitation della singola CVE	Modello predittivo su dati aggregati (EPSS) [11], [20]	Una probabilità per ogni CVE pubblicata	Dati pubblici verificabili	Telemetria dei partner rappresentativa del contesto
Severità, categoria e prodotti di una CVE non arricchita	Classificazione dal testo, LLM [21], [22]	Attributi stimati dalla descrizione	Studi peer-reviewed e benchmark; verifica su dati pubblici in questo articolo	Descrizioni simili a quelle di addestramento; verifica a campione
Eventi descritti solo in forma narrativa	Estrazione con NLP e LLM [22]	Serie di eventi classificati	Benchmark, con errori documentati	Controllo umano sui dati estratti
Pochi eventi interni	Aggiornamento bayesiano, Bayesian network [9], [24]	Stima che combina fonti esterne e locali, con il relativo intervallo	Metodologica; verifica su dati pubblici in questo articolo	Stima a priori dichiarata e pertinente
Classi rare nei dati di addestramento	Dati sintetici [25]	Volume per addestrare e collaudare i modelli	Rassegna critica	Nessuna pretesa di nuova informazione
Perdite non dichiarate, eventi estremi	Nessuna tecnica risolutiva [19]	Scenari fondati su ipotesi	Teoria dei dati mancanti	Ipotesi esplicite e analisi di sensibilità

06 Evidenze disponibili

Il volume dei dati da trattare

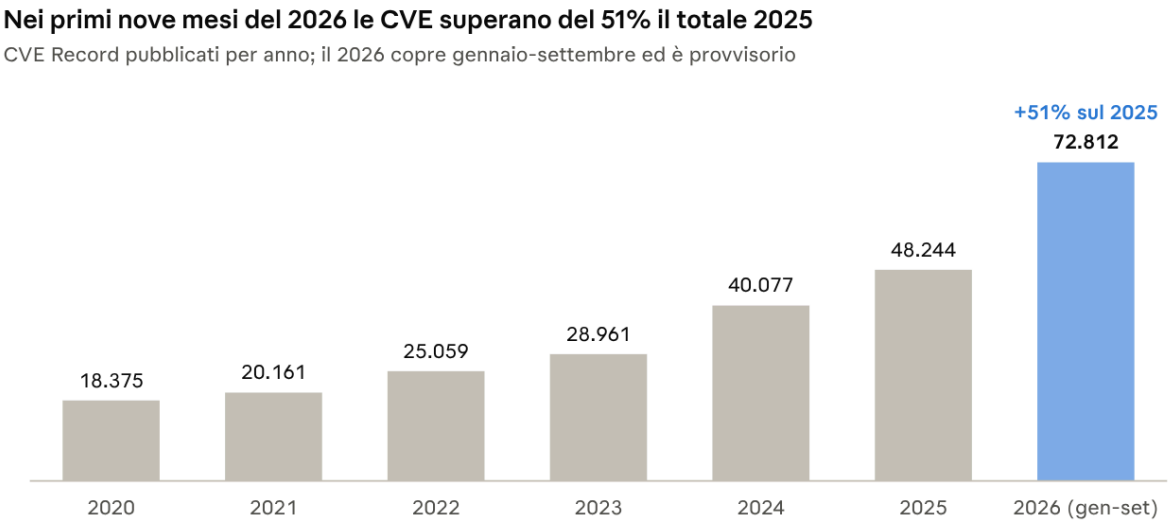


Grafico 1. CVE Record pubblicati per anno, 2020-2025 e gennaio-settembre 2026. Fonte: CVE Program, Metrics [6].

In cinque anni il numero di CVE pubblicate in dodici mesi è cresciuto di oltre due volte e mezzo, e nei primi tre trimestri del 2026 ha superato di metà il totale dell'anno precedente [6]; il dato del 2026 è parziale e provvisorio. Il NIST riferisce a sua volta che nel primo trimestre del 2026 le CVE ricevute sono state quasi un terzo in più rispetto allo stesso periodo del 2025 [5]. Poiché l'arricchimento da parte dell'ente è ormai limitato alle categorie prioritarie, il numero dei record privi di attributi cresce insieme al totale, e con esso la parte del modello quantitativo che deve essere alimentata da stime.

La qualità delle stime sostitutive

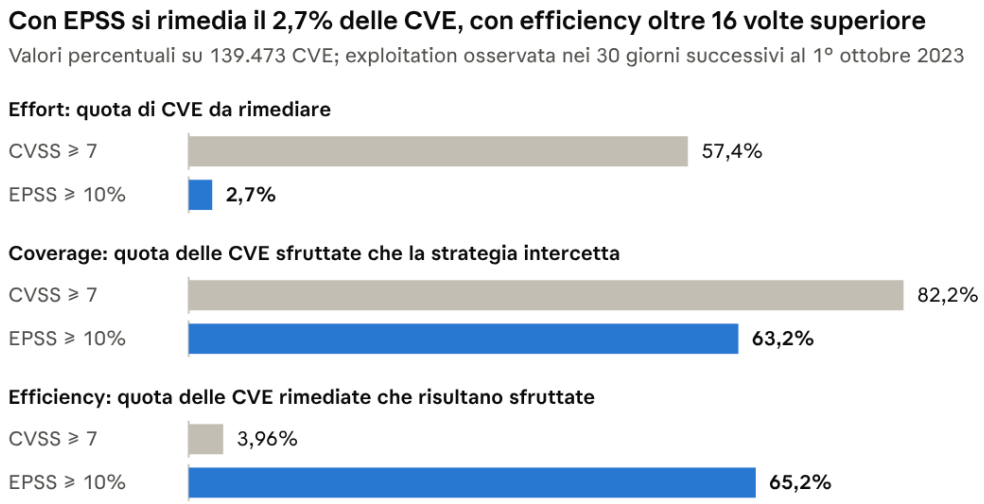


Grafico 2. Strategie di remediation CVSS ≥ 7 ed EPSS ≥ 10% sulla stessa popolazione di CVE. Fonte: FIRST, The EPSS Model [11].

Il confronto pubblicato da FIRST consente di valutare quanto una probabilità stimata sia utile in luogo di un dato assente. Rimediare tutte le vulnerabilità con CVSS pari o superiore a 7 comporta intervenire su 80.024 CVE per intercettarne 3.166 con exploitation osservata nel mese successivo; con una soglia EPSS del 10% le CVE da trattare scendono a 3.735 e quelle intercettate sono 2.435 [11]. Lo sforzo si riduce di oltre venti volte e la coverage diminuisce di diciannove punti percentuali. FIRST precisa che il 10% non è una soglia ufficiale. Abbassandola si recupera coverage a costo di più lavoro, e la scelta del punto di equilibrio resta una decisione di risk appetite.

Per le altre modalità di intervento le prove sono di natura diversa. Il completamento degli attributi e l'estrazione dai testi sono documentati da studi sperimentali e da benchmark [21], [22], che misurano l'accuratezza del modello e non l'effetto sulla stima finale del rischio. Tra le fonti che ho esaminato non compaiono studi che confrontino, presso organizzazioni reali, le stime quantitative ottenute con e senza queste tecniche.

I parametri pubblici per la calibrazione

La Tabella 5 raccoglie i valori pubblicati nel 2026 che possono servire da riferimento esterno nella stima di frequenza e magnitudo.

Tabella 5. Parametri pubblici utilizzabili come riferimento esterno.

Parametro	Valore	Fattore del modello	Fonte
Costo medio globale di un data breach	4,99 milioni di dollari (+12% sull'anno precedente)	Loss Magnitude	[10]
Costo medio di un breach malevolo AI-enabled	6 milioni di dollari	Loss Magnitude	[10]

Parametro	Valore	Fattore del modello	Fonte
Differenziale di costo associato all'uso estensivo di AI e automation nella security, rispetto a nessun uso	1,93 milioni di dollari in meno per breach	Effetto dei controlli	[10]
Breach con sfruttamento di vulnerabilità come vettore iniziale	31%	Threat Event Frequency	[7]
Breach con coinvolgimento di terze parti	48%	Threat Event Frequency	[7]
CVE con exploitation osservata entro 30 giorni	3.852 su 139.473	Threat Event Frequency	[11]
Vulnerabilità KEV rimediate completamente	26% (38% l'anno precedente)	Vulnerability	[7]
Tempo mediano di risoluzione completa delle vulnerabilità KEV	43 giorni (32 l'anno precedente)	Vulnerability	[7]

Questi valori sono adatti a costruire stime a priori e non misurano la situazione della singola organizzazione. Il differenziale di 1,93 milioni, in particolare, nasce dal confronto tra gruppi di organizzazioni che hanno scelto se adottare o meno gli strumenti. È plausibile che chi li adotta in modo estensivo disponga anche di budget, processi e competenze superiori, e i dati pubblicati non permettono di separare l'effetto della tecnologia da quello della maturità complessiva. Impiegarlo come effetto causale di un investimento porterebbe a sovrastimarne il beneficio.

07 Un esempio: stimare il rischio con un solo evento osservato

L'esempio è illustrativo. L'organizzazione, che chiamo Alfa, è ipotetica, e i parametri presentati come ipotesi sono scelte mie. Voglio mostrare come cambia una stima quantitativa quando alle poche osservazioni interne si affiancano dati esterni ottenuti con le tecniche descritte.

Il dato che Alfa non possiede

Alfa è un'impresa di servizi con circa 1.500 dipendenti, soggetto importante ai sensi della NIS2. Il vulnerability scan restituisce 10.000 CVE distinte ancora aperte e il team ne chiude circa 400 al mese. Alfa non ha elementi propri per stabilire quali saranno sfruttate. Assumendo che il backlog abbia la stessa composizione della popolazione analizzata da FIRST [11], le alternative sono quelle della Tabella 6.

Tabella 6. Due criteri di ordinamento applicati al backlog di Alfa. Proporzioni da [11]; backlog e capacità sono ipotesi.

Grandezza	CVSS pari o superiore a 7	EPSS pari o superiore al 10%
CVE da rimediare	5.738	268
Di cui con exploitation attesa entro 30 giorni	227	175
Tempo per esaurire la coda	14,3 mesi	circa 20 giorni
CVE con exploitation attesa chiuse nel primo mese	circa 16 su 276	175 su 276

La probabilità fornita da EPSS sostituisce un'informazione che Alfa non potrebbe produrre e permette di chiudere nel primo mese 175 delle 276 vulnerabilità per le quali è attesa attività di exploitation, contro le 16 circa che si otterrebbero procedendo per severità senza altro criterio. Il punteggio descrive l'exploitation osservata a livello globale e non considera l'impatto sul singolo asset; va quindi combinato con la criticità dei sistemi interessati.

La frequenza degli eventi

Considero lo scenario «data breach per exploitation di una vulnerabilità su un sistema esposto». Nei sei anni per i quali Alfa conserva un registro degli incidenti l'evento si è verificato una volta. Tratto la frequenza annua λ come grandezza incerta, con distribuzione a priori Gamma e verosimiglianza di Poisson, cosicché dopo k eventi osservati in T anni la distribuzione a posteriori è ancora una Gamma:

$$\lambda \mid k, T \sim \text{Gamma}\left(\alpha + k,\ \beta + T\right)$$

Confronto tre basi informative. La prima usa i soli dati locali, con una distribuzione a priori non informativa ($\alpha = 0{,}5$ e $\beta = 0$). La seconda usa soltanto una stima esterna, che suppongo ricavata da dati aggregati e adattata al profilo di esposizione di Alfa: media di 0,30 eventi per anno e contenuto informativo pari a quattro eventi osservati in circa tredici anni ($\alpha = 4$ e $\beta = 13{,}3$). La terza aggiorna la stima esterna con l'osservazione locale. Il Grafico 3 riporta la media e l'intervallo di credibilità al 90% nei tre casi.

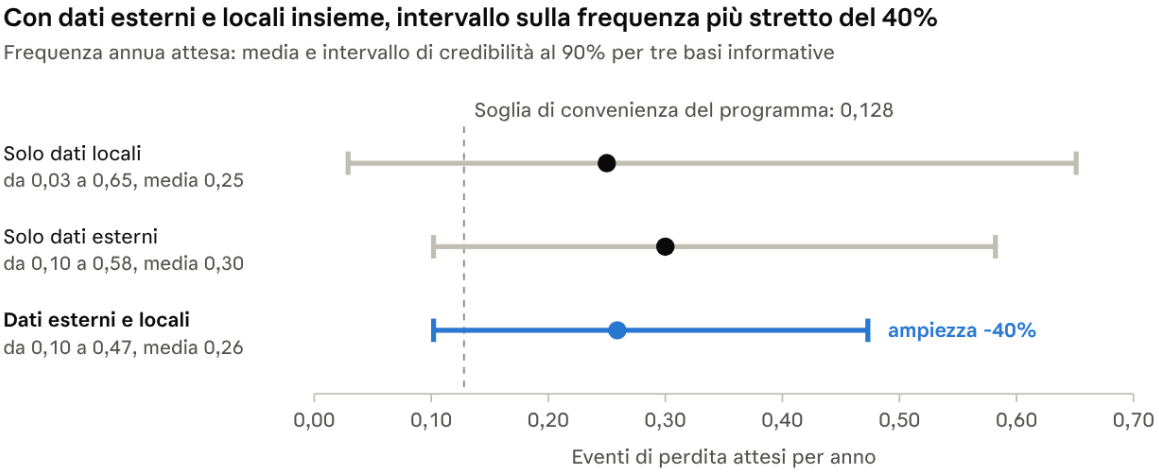


Grafico 3. Frequenza annua dello scenario per tre basi informative. Calcolo dell'autore su ipotesi illustrative.

Con i soli dati locali l'intervallo va da 0,03 a 0,65 eventi per anno. L'estremo superiore è oltre venti volte quello inferiore, e una stima così dispersa serve a poco in sede di decisione. La combinazione con la stima esterna restringe l'intervallo tra 0,10 e 0,47, con un'ampiezza inferiore del 40%, e lascia quasi invariata la media.

Dalla stima alla decisione

Per la magnitudo Alfa dispone di una sola osservazione, insufficiente a stimare una distribuzione. Adotto perciò in tutti i casi una lognormale con mediana di 1,2 milioni di euro e 90° percentile di 6 milioni, che suppongo calibrata su dati esterni; ne risulta una perdita media per evento, $E[L]$, di 2,64 milioni. Alfa valuta un programma di remediation prioritizzata e di detection dal costo C di 180.000 euro l'anno, dal quale si attende una riduzione di un terzo della frequenza e del 30% della magnitudo. Sotto queste ipotesi la perdita annua attesa (ALE) si riduce del 53%, e il programma si ripaga quando la frequenza supera la soglia

$$\lambda^{*} = \frac{C}{(1 - r_f\, r_m)\; E[L]} = 0{,}128$$

dove $r_f = 2/3$ e $r_m = 0{,}7$ sono i fattori di riduzione ipotizzati per frequenza e magnitudo.

Tabella 7. Risultati con i soli dati locali e con dati locali ed esterni combinati. Importi in euro.

Indicatore	Solo dati locali	Dati locali ed esterni
Frequenza annua attesa	0,25	0,26
Intervallo di credibilità al 90% sulla frequenza	da 0,03 a 0,65	da 0,10 a 0,47
Perdita annua attesa (ALE)	660.000	683.000
Intervallo al 90% sulla ALE	da 77.000 a 1.720.000	da 269.000 a 1.250.000
Probabilità che il programma si ripaghi	67,4%	89,5%
Beneficio netto atteso del programma	172.000	184.000

Le due basi informative conducono quasi alla stessa perdita attesa, e chi guardasse soltanto la media non vedrebbe differenze. Cambia il grado di fiducia con cui Alfa può decidere: con i soli dati locali un terzo della distribuzione della frequenza si trova sotto la soglia di convenienza, con i dati combinati circa un decimo. Ho simulato la perdita annua con il metodo Monte Carlo sulla stima combinata, con due milioni di iterazioni: il 95° percentile è di 3,73 milioni di euro senza il programma e di 1,70 milioni con il programma.

L'esempio ha tre limiti. La stima esterna è un'ipotesi, e una stima a priori errata sposta il risultato tanto più quanto meno numerosi sono i dati locali. L'effetto del programma è anch'esso ipotizzato, per le ragioni esposte a proposito dell'efficacia dei controlli. La distribuzione della magnitudo non tiene conto delle perdite non dichiarate, che il campione esterno per costruzione non contiene. Nella sezione che segue misuro su dati reali quanto pesino il primo e il terzo.

08 Una verifica su dati pubblici

L'esempio di Alfa poggia su ipotesi dichiarate: una stima esterna con un contenuto informativo fissato, una frequenza costante nel tempo, una perdita lognormale. Per valutare quanto reggano, ho misurato le stesse grandezze su dati reali. La frequenza degli incidenti e le perdite di una singola organizzazione non sono pubbliche. Lo sono, con origine e data certe, i dati sulle vulnerabilità e sul loro sfruttamento e una raccolta di incidenti codificati (Tabella 8). La verifica riguarda quindi un dominio vicino a quello di interesse, in cui il valore da stimare è osservabile. Comprende cinque prove: la misura delle lacune, il completamento dei punteggi di severità, la stima di un tasso con poche osservazioni, l'effetto dei dati non dichiarati su severità reali e il confronto tra EPSS e CVSS rispetto a un esito pubblico.

Tabella 8. Fonti dei dati della verifica.

Fonte	Contenuto	Versione
National Vulnerability Database, letta da un mirror pubblico [26]	332.001 CVE non respinte, pubblicate dal 3 gennaio 2016 al 21 novembre 2026	Stato del 21 novembre 2026
Catalogo Known Exploited Vulnerabilities di CISA [27]	1.790 vulnerabilità di cui è noto lo sfruttamento	Versione 2026.11.20
Punteggi EPSS, letti da un aggregatore pubblico [28]	Punteggio di ogni CVE a una data	Istantanee del 10 aprile e del 13 luglio 2025
VERIS Community Database [29]	10.047 incidenti codificati nello schema VERIS	Stato del 20 novembre 2026

Ogni prova usa un esito osservabile da chiunque. Valuto i modelli su dati successivi a quelli di addestramento, oppure su campioni estratti da una popolazione di cui è noto il valore reale. I punteggi EPSS provengono da un aggregatore di terzi, perché l'archivio ufficiale non era raggiungibile dal mio ambiente di calcolo. Ho confrontato con l'istantanea venti valori ufficiali del 10 aprile 2025: delle 17 CVE presenti, 16 avevano valore identico e una uno scarto di 0,00005.

Le lacune misurate

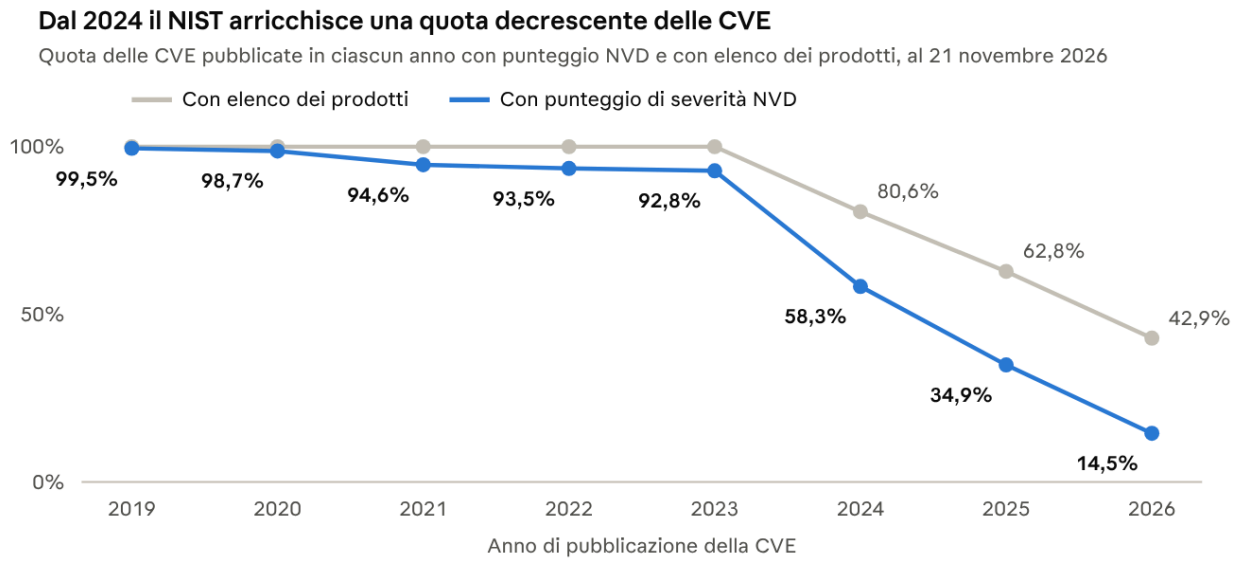


Grafico 4. Quota delle CVE con punteggio NVD e con elenco dei prodotti, per anno di pubblicazione. Elaborazione dell'autore sui dati NVD al 21 novembre 2026 [26].

Fino al 2023 oltre il 92% delle CVE pubblicate in un anno riceveva un punteggio di severità dal NIST, e tutte ricevevano l'elenco dei prodotti interessati (Grafico 4). La quota con punteggio NVD scende al 58,3% per le CVE del 2024, al 34,9% per quelle del 2025 e al 14,5% per quelle del 2026; la quota con elenco dei prodotti scende all'80,6%, al 62,8% e al 42,9%. Tra le 71.569 CVE pubblicate dal 15 aprile al 31 ottobre 2026, dopo la decisione del NIST [5], l'11,0% ha un punteggio NVD e il 41,4% l'elenco dei prodotti. Nello stesso periodo del 2025 le quote erano il 38,3% e il 65,9%. Sulla quota del punteggio del 2026 pesa anche la rinuncia al punteggio duplicato annunciata dal NIST [5], mentre il calo del 2024 e del 2025, anteriore a quella decisione, è coerente con l'arretrato accumulato dall'ente. Come misura dell'analisi svolta dal NIST considero quindi più fedele l'elenco dei prodotti.

Una CVE priva del punteggio NVD ha di norma quello di un'altra fonte. Nel periodo successivo alla decisione l'85,0% delle CVE ha soltanto un punteggio di questo tipo, in prevalenza quello della CNA che ha pubblicato la vulnerabilità, e il 3,9% non ne ha alcuno. Il dato che manca più spesso è l'elenco dei prodotti in formato Common Platform Enumeration (CPE), su cui poggia l'associazione automatica della vulnerabilità agli asset dell'organizzazione. Le 151 CVE di quel periodo già entrate nel catalogo KEV hanno tutte l'elenco dei prodotti, mentre il punteggio NVD compare nel 29,8% dei casi. Per le perdite la lacuna è più ampia: nel VERIS Community Database 324 incidenti su 10.047, il 3,2%, riportano un importo complessivo della perdita, e 5.127, il 51,0%, il numero di record compromessi.

Il completamento dei punteggi di severità

Questa prova riguarda la seconda modalità di intervento, il completamento dei record parziali. Il modello riceve la descrizione testuale di una CVE e ne predice le otto metriche base del vettore CVSS v3.1; calcolo poi il punteggio con la formula della specifica [30], in un'implementazione che riproduce il punteggio registrato nella NVD per tutti i 153.044 vettori assegnati dal NIST. Si tratta di un classificatore lineare a regressione logistica su pesi TF-IDF, costruito con la libreria scikit-learn

[31] e addestrato sulle 105.134 CVE pubblicate dal 2017 al 2022 con vettore assegnato dal NIST. L'ho scelto volutamente semplice: non rappresenta lo stato dell'arte, e i suoi risultati vanno letti come un termine di riferimento di base. Lo valuto sulle CVE pubblicate dal 2023 al 2026, che il modello non ha mai visto.

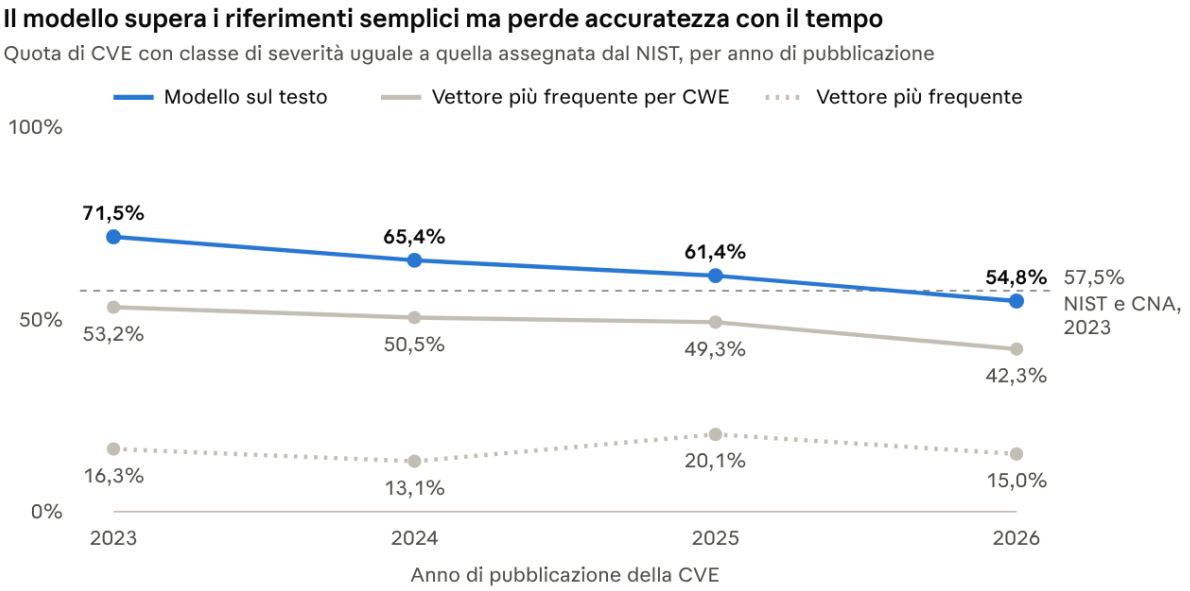


Grafico 5. CVE con la stessa classe di severità assegnata dal NIST: modello sul testo e due riferimenti semplici, per anno di pubblicazione. La linea tratteggiata indica l'accordo tra NIST e CNA nel 2023. Elaborazione dell'autore sui dati NVD [26].

Sulle CVE del 2023 il modello assegna la stessa classe di severità del NIST nel 71,5% dei casi, con un errore medio di 0,80 punti, contro il 53,2% di un riferimento che attribuisce a ogni CVE il vettore più frequente della sua categoria CWE e il 16,3% di uno che attribuisce a tutte il vettore più frequente nei dati di addestramento (Grafico 5). Ritengo più informativo un altro termine di confronto, l'accordo tra due valutazioni umane della stessa vulnerabilità. Sulle 15.003 CVE del 2023 che hanno sia il punteggio del NIST sia quello della CNA, il secondo cade nella stessa classe del primo nel 57,5% dei casi; sulle stesse CVE il modello concorda con il NIST nel 70,4%.

L'accuratezza diminuisce con la distanza dal periodo di addestramento. In tre anni la quota di classi corrette passa dal 71,5% al 54,8% e l'errore medio da 0,80 a 1,25 punti. L'intervallo al 90% ricavato dalle probabilità del modello resta calibrato, con una copertura tra il 91,5% e il 93,6%, e si allarga da 3,3 a 4,5 punti. Per le 3.480 CVE del 2026 prive di qualsiasi punteggio l'intervallo è ampio in media 6,1 punti su una scala di dieci: la sola descrizione non basta a determinare la severità. Un punteggio completato va quindi trattato come una distribuzione, e il modello va riaddestrato con regolarità.

La stima di un tasso con poche osservazioni

Questa prova ripete su dati reali il confronto dell'esempio. La grandezza da stimare è il tasso di sfruttamento delle vulnerabilità di un vendor, cioè la quota delle sue CVE che entra nel catalogo KEV. Il tasso svolge il ruolo della frequenza degli eventi, e il numero di CVE osservate quello degli anni di osservazione. A differenza della frequenza di Alfa, il valore reale è noto: è la quota calcolata

su tutte le CVE del vendor. Le CVE considerate sono le 168.868 pubblicate dal 2017 al 2024 a cui la NVD associa un vendor; 1.109 sono nel catalogo KEV, lo 0,66%. Valuto i 26 vendor con almeno 1.000 CVE, i cui tassi vanno da zero al 3,08%.

In questa prova la stima esterna è ricavata dai dati: è la distribuzione Gamma che descrive i tassi degli altri vendor con almeno 50 CVE, che stimo a massima verosimiglianza escludendo ogni volta il vendor valutato. Per ogni vendor considero l'estrazione casuale di n CVE e calcolo in forma esatta, sulla distribuzione ipergeometrica del numero di CVE sfruttate nel campione, l'errore quadratico medio della stima, l'ampiezza dell'intervallo di credibilità al 90% e la sua copertura, cioè la quota dei casi in cui l'intervallo contiene il valore reale (Tabella 9).

Tabella 9. Stima del tasso di sfruttamento da un campione di n CVE. Media sui 26 vendor; errore e ampiezza in punti percentuali.

CVE osservate	Errore, soli dati locali	Errore, stima combinata	Ampiezza, soli dati locali	Ampiezza, stima combinata	Copertura, soli dati locali	Copertura, stima combinata
10	5,72	0,85	20,6	2,9	91,5%	92,9%
25	2,66	0,85	9,0	2,6	89,4%	86,7%
50	1,59	0,79	5,1	2,2	90,4%	88,9%
100	1,00	0,67	3,1	1,9	88,3%	88,0%
250	0,56	0,47	1,8	1,4	89,8%	84,5%
500	0,35	0,32	1,2	1,1	91,3%	90,3%

Con 50 CVE osservate la combinazione dimezza l'errore, da 1,59 a 0,79 punti percentuali, e riduce del 57% l'ampiezza dell'intervallo, con una copertura dell'88,9% contro il 90,4% dei soli dati locali. Ai soli dati locali occorrono in mediana 180 CVE per ottenere l'ampiezza che la stima combinata raggiunge con 50. Il beneficio si esaurisce al crescere dei dati: con 500 CVE i due metodi si equivalgono.

Il beneficio dipende dalla pertinenza della stima esterna, e in una sola direzione. Per i quattro vendor il cui tasso reale è almeno triplo della media esterna, l'intervallo combinato costruito su 25 CVE contiene il valore reale nel 43,2% dei casi contro il 90% dichiarato, e servono 100 CVE perché la copertura torni all'85%. Quando la stima esterna sopravvaluta il tasso di almeno tre volte, la copertura della stima combinata resta superiore a quella dei soli dati locali e l'errore è da due a cinque volte minore.

Due ipotesi dell'esempio non trovano riscontro. La prima riguarda il contenuto informativo della stima esterna. Il parametro di forma della distribuzione appresa dai dati, che misura a quanti eventi osservati equivale la stima, è 0,23 contro il 4 ipotizzato per Alfa: l'eterogeneità tra le unità è tale che la media della popolazione dice poco sulla singola unità. Se la dispersione delle frequenze tra organizzazioni fosse quella osservata tra i vendor, la stima esterna di Alfa varrebbe meno di un anno di osservazione, l'intervallo al 90% sulla frequenza andrebbe da 0,015 a 0,507 e la probabilità che il programma si ripaghi scenderebbe dall'89,5% al 52,3%. La trasposizione è un'ipotesi, perché riguarda popolazioni diverse. Indica tuttavia che la conclusione dell'esempio dipende dal contenuto informativo attribuito alla stima esterna più che dalla sua media.

La seconda riguarda la stazionarietà. Ho usato i tassi stimati sulle CVE pubblicate dal 2017 al 2020 per prevedere quante delle CVE pubblicate dal 2021 al 2024 sarebbero entrate nel catalogo, per i 29 vendor con almeno 200 CVE in entrambi i periodi. Il tasso complessivo cambia poco, dallo 0,69% allo 0,64%; quello dei singoli vendor varia di oltre un fattore 2 nel 62% dei casi. Gli intervalli predittivi al 90% costruiti sull'intera storia contengono il valore poi osservato nel 55,2% dei casi con la stima combinata e nel 62,1% con i soli dati locali. Riducendo a un decimo il peso delle osservazioni del primo periodo, la copertura della stima combinata risale al 93,1%. Ho scelto il fattore esplorando gli stessi dati su cui lo valuto, e va quindi preso come indicazione di un ordine di grandezza.

Dati non dichiarati e code pesanti

Il meccanismo per cui un incidente non viene reso pubblico non è osservabile nei dati, e resta quindi simulato; le severità sono reali. Per i 5.127 incidenti del VERIS Community Database di cui è noto il numero di record compromessi suppongo dichiarato il 30% dei casi, con probabilità che dipende da 45 variabili osservate (classe dimensionale, settore, periodo, tipo di azione e di attore, paese) oppure dalla severità stessa. Stimo la severità tipica, cioè la media geometrica del numero di record, e faccio variare l'intensità della selezione su quattro livelli.

Quando la selezione dipende dalle variabili osservate, i soli casi dichiarati sottostimano la severità tipica dal 67% al 95%, e un'imputazione fondata sulle stesse variabili riduce lo scarto a valori compresi tra il 3% e il 34%. Quando dipende dalla severità, la sottostima va dall'80% al 99% sui soli casi dichiarati e dal 61% al 96% dopo l'imputazione. È la distinzione tra i meccanismi di mancanza [19], misurata su severità reali: il completamento recupera gran parte della distorsione nel primo caso e una parte minore nel secondo.

Le severità reali sono anche molto più disperse di quelle ipotizzate per Alfa. La deviazione standard del logaritmo è 4,9 contro 1,26; la mediana è di 1.112 record e la media di 14,5 milioni; l'1% degli incidenti più gravi contiene il 90,6% dei record. L'indice di coda, stimato con lo stimatore di Hill sulle 100 osservazioni più grandi, è 0,77, vicino allo 0,7 trovato da Maillart e Sornette su un altro insieme di dati [32]. Con un indice inferiore a 1 la media della distribuzione teorica non è finita: la perdita media non è stimabile in modo stabile, e l'analisi deve poggiare su quantili e scenari.

EPSS e CVSS rispetto a un esito pubblico

Il confronto del Grafico 2 usa l'attività di exploitation rilevata dai sensori dei partner di FIRST, che non è pubblica. Qui uso come esito l'inserimento della CVE nel catalogo KEV nei dodici mesi successivi alla data dei punteggi. La popolazione è formata dalle CVE pubblicate dal 2016, dotate di punteggio EPSS a quella data e non ancora presenti nel catalogo: 194.474 al 10 aprile 2025 e 206.108 al 13 luglio 2025. Gli inserimenti successivi sono 64 e 71. Il punteggio CVSS è quello del NIST o, in sua assenza, quello di un'altra fonte. Nella Tabella 10 l'effort è la quota delle CVE da trattare, la coverage la quota degli inserimenti intercettati e l'efficiency la quota delle CVE trattate che entrano nel catalogo.

Tabella 10. Strategie di remediation valutate sull'inserimento nel catalogo KEV nei dodici mesi successivi alla data dei punteggi.

Strategia	Effort, 10 aprile 2025	Coverage	Efficiency	Effort, 13 luglio 2025	Coverage	Efficiency
CVSS ≥ 7	54,3%	87,5%	0,053%	54,0%	90,1%	0,057%
CVSS ≥ 9	14,1%	39,1%	0,091%	14,1%	29,6%	0,072%
EPSS ≥ 10%	5,2%	50,0%	0,317%	4,8%	45,1%	0,324%
EPSS, stesso effort di CVSS ≥ 9	14,1%	60,9%	0,143%	14,1%	57,7%	0,141%
EPSS, stesso effort di CVSS ≥ 7	54,3%	78,1%	0,047%	54,0%	77,5%	0,049%

La soglia EPSS del 10% richiede un decimo del lavoro della soglia CVSS 7 e ha un'efficiency circa sei volte maggiore, ma copre circa la metà degli inserimenti successivi, il 50,0% e il 45,1%, contro l'87,5% e il 90,1%. A parità di effort l'esito dipende dal livello: con il 14% delle CVE, EPSS copre circa il 60% degli inserimenti e CVSS ≥ 9 tra il 30% e il 39%; con il 54%, CVSS ≥ 7 copre più di EPSS. Il vantaggio misurato da FIRST è quindi confermato nel verso e ridimensionato nell'entità. Il rapporto tra le efficiency è 6, contro il 16 del confronto di FIRST, perché l'inserimento nel catalogo è un evento molto più raro dell'attività rilevata dai sensori. Una graduatoria calcolata a una data ha anche un limite che nessun punteggio rimuove: dei 240 inserimenti nel catalogo nei dodici mesi successivi al 10 aprile 2025, 165 riguardano CVE pubblicate dopo quella data.

Che cosa regge e che cosa no

Tabella 11. Affermazioni e ipotesi dell'esempio a confronto con i dati.

Affermazione o ipotesi	Riscontro sui dati	Esito
La combinazione con una stima esterna riduce l'incertezza quando i dati locali sono pochi	Con 50 CVE errore dimezzato e ampiezza dell'intervallo ridotta del 57%	Confermato
Una stima esterna errata sposta il risultato	Copertura del 43,2% con 25 CVE quando la stima esterna sottovaluta di almeno tre volte; copertura non inferiore a quella dei soli dati locali quando sopravvaluta	Confermato in una direzione
La stima esterna vale quattro eventi osservati	Appresa dai dati vale 0,23 eventi; la probabilità che il programma di Alfa si ripaghi scende dall'89,5% al 52,3%	Non confermato
La frequenza è costante nel tempo	Con l'intera storia la copertura predittiva è del 55-62%	Non confermato
La perdita per evento è lognormale	Indice di coda 0,77; l'1% degli incidenti contiene il 90,6% dei record	Non confermato
EPSS concentra la remediation meglio di CVSS	Efficiency circa sei volte maggiore con un decimo dell'effort; coverage quasi dimezzata	Confermato in parte
Un modello può completare i punteggi di severità mancanti	Classe corretta nel 71,5% dei casi nel primo anno e nel 54,8% nel quarto	Confermato, con degrado nel tempo

I meccanismi descritti reggono alla prova dei dati (Tabella 11). La combinazione di fonti riduce l'incertezza, una stima esterna non pertinente la falsa, l'imputazione rimedia alla selezione su variabili osservate assai più che a quella sul valore. Non reggono tre ipotesi dell'esempio: il contenuto informativo della stima esterna, la stazionarietà e la forma lognormale delle perdite. Le prime due portano a sopravvalutare la precisione delle stime, la terza a sottovalutare il peso degli eventi estremi.

La verifica ha limiti propri. Usa grandezze vicine a quelle di interesse: l'unità è il vendor, mentre l'analisi del rischio riguarda l'organizzazione, e l'evento è l'inserimento di una vulnerabilità nel catalogo KEV, che di per sé non comporta una perdita. L'inserimento dipende dalle scelte di chi cura il catalogo e registra lo sfruttamento con ritardo. Il meccanismo di selezione sulle severità resta simulato. Nel confronto tra EPSS e CVSS i punteggi CVSS sono quelli disponibili a novembre 2026, che possono differire da quelli noti alla data delle istantanee. Il confronto delle stime con le perdite effettive di un'organizzazione resta da compiere.

09 Vincoli normativi e standard

Le norme europee non prescrivono l'uso dell'AI né l'adozione di un metodo quantitativo. Alcune fissano requisiti che presuppongono una misura del rischio; altre disciplinano i sistemi di AI impiegati per ottenerla.

La direttiva NIS2 richiede misure tecniche, operative e organizzative adeguate ai rischi, definite tenendo conto dello stato dell'arte, degli standard pertinenti e dei costi di attuazione. La proporzionalità va valutata in base all'esposizione al rischio, alla dimensione del soggetto e alla probabilità e gravità degli incidenti. Tra le misure minime figurano le politiche di analisi dei rischi e le procedure per valutare l'efficacia delle misure adottate (art. 21). Gli organi di gestione approvano le misure e ne sovrintendono l'attuazione (art. 20). Gli incidenti significativi vanno segnalati con un preallarme entro 24 ore, una notifica entro 72 ore e una relazione finale entro un mese dalla notifica (art. 23) [1].

In Italia la direttiva è stata recepita con il d.lgs. 138/2024. Le determinazioni dell'Agenzia per la cybersicurezza nazionale hanno fissato, per i soggetti inseriti nell'elenco NIS nel 2025, un termine di 18 mesi dalla comunicazione di inserimento per l'adozione delle misure di sicurezza di base e di 9 mesi per l'obbligo di notifica; per i soggetti inseriti nel 2026 le misure vanno adottate entro il 31 luglio 2027 e la notifica è dovuta dal 1° gennaio 2027 [33].

Il regolamento DORA, applicabile dal 17 gennaio 2025, chiede alle entità finanziarie un framework documentato di ICT risk management sotto la responsabilità dell'organo di gestione, la rilevazione tempestiva delle anomalie, il monitoraggio continuo della sicurezza dei sistemi ICT e la gestione del rischio derivante da terze parti [2]. Le notifiche previste dalle due norme sono rivolte alle autorità competenti e non al pubblico, e per questa ragione non riducono di per sé la scarsità di dati di cui dispone chi deve stimare il rischio [4].

L'AI Act stabilisce, per i sistemi ad alto rischio, requisiti di accuracy, robustness e cybersecurity lungo l'intero ciclo di vita (art. 15). Il regolamento (UE) 2026/1744, in vigore dal 27 luglio 2026, ha differito l'applicazione degli obblighi al 2 dicembre 2027 per i sistemi ad alto rischio dell'Allegato III e al 2 agosto 2028 per quelli incorporati nei prodotti dell'Allegato I [34]. Se uno strumento di AI impiegato per l'analisi del rischio cyber rientri tra i sistemi ad alto rischio dipende dall'uso concreto e va verificato caso per caso; i criteri dell'art. 15 offrono comunque un riferimento per la valutazione dei fornitori. Tra gli standard volontari, il NIST AI Risk Management Framework articola la gestione dei rischi dell'AI nelle funzioni Govern, Map, Measure e Manage [35], e la norma ISO/IEC 42001:2023 specifica i requisiti di un sistema di gestione dell'AI [36].

Tabella 12. Riferimenti normativi e standard, con la rilevanza per l'analisi quantitativa.

Riferimento	Natura e ambito	Requisito rilevante	Rilevanza per l'analisi quantitativa
NIS2, direttiva (UE) 2022/2555 [1]	Direttiva; soggetti essenziali e importanti	Misure proporzionate a probabilità e gravità degli incidenti (art. 21); approvazione degli organi di gestione (art. 20)	Richiede valutazioni motivate di probabilità e di impatto

Riferimento	Natura e ambito	Requisito rilevante	Rilevanza per l'analisi quantitativa
D.lgs. 138/2024 e determinazioni ACN [33]	Recepimento italiano	Misure di sicurezza di base, notifica degli incidenti significativi	Scadenze per l'adozione delle misure
DORA, regolamento (UE) 2022/2554 [2]	Regolamento; settore finanziario	Framework di ICT risk management, monitoraggio continuo, terze parti	Richiede una misura del rischio aggiornata
AI Act, regolamento (UE) 2024/1689 [34]	Regolamento; sistemi di AI	Accuracy, robustness e cybersecurity (art. 15)	Requisiti per i modelli che producono le stime
ISO/IEC 27005:2022 [12]	Standard internazionale	Ciclo di gestione del rischio	Processo entro cui si colloca la quantificazione
Open FAIR [16]	Standard di The Open Group	Tassonomia e metodo di analisi	Struttura del modello quantitativo
NIST AI RMF 1.0 [35] e ISO/IEC 42001:2023 [36]	Framework e standard volontari	Governo dei rischi dei sistemi di AI	Governo dei modelli impiegati

10 Ciò che l'AI non ricostruisce

Dati mancanti per ragioni legate al loro valore

Il limite principale discende dalla distinzione tra i meccanismi di mancanza [19]. Le tecniche di completamento stimano un valore assente a partire da ciò che si osserva, e funzionano quando l'assenza dipende da caratteristiche note. Le perdite e gli incidenti che un'organizzazione preferisce non divulgare possono mancare per ragioni legate alla loro entità. Un modello addestrato sui casi pubblici apprende la distribuzione dei casi pubblici: se le perdite taciute sono le più alte, le stime che ne derivano sono distorte verso il basso in una misura che i dati disponibili non permettono di calcolare. La prova su severità reali ne indica l'ordine di grandezza sotto un meccanismo ipotizzato: con il 30% dei casi dichiarati e una selezione legata alla severità, il valore tipico risulta sottostimato dall'80% al 99%.

Lo stesso vale per gli eventi estremi, che Eling e Wirfs separano da quelli ordinari con un metodo della extreme value theory [18] e che i campioni disponibili osservano poche volte. Nella raccolta di incidenti che ho esaminato l'1% dei casi contiene il 90,6% dei record compromessi. I dati sintetici non modificano questa situazione [25]. In questi casi l'analisi quantitativa resta praticabile con ipotesi esplicite sul meccanismo di mancanza e con un'analisi di sensibilità che mostri quanto il risultato ne dipenda.

Validità dei modelli

Arp e colleghi hanno individuato dieci errori ricorrenti nella progettazione e nella valutazione dei sistemi di sicurezza basati su machine learning e ne hanno riscontrato la diffusione in 30 articoli pubblicati in dieci anni nelle principali conferenze del settore [37]. Sommer e Paxson avevano già osservato, a proposito dell'anomaly detection, la distanza tra i risultati della ricerca e l'impiego operativo [38].

Per i modelli che completano gli attributi delle vulnerabilità si aggiunge una difficoltà specifica. Dal 2026 i record arricchiti dal NIST appartengono alle categorie prioritarie, mentre quelli da completare appartengono alle altre, per le quali resta al più il punteggio attribuito da chi ha assegnato la CVE [5]. La popolazione su cui il modello viene addestrato può quindi differire da quella a cui viene applicato, e l'accuratezza misurata sulla prima non si trasferisce automaticamente alla seconda. Nella verifica su dati pubblici la quota di classi corrette scende di quasi 17 punti in tre anni.

Errori dei modelli generativi

Gli autori di CTIBench ricordano che su affidabilità, accuratezza e hallucination dei large language model nei compiti di threat intelligence restano riserve [22]. Un attributo stimato o un evento estratto da un testo entrano nel modello quantitativo con l'aspetto di un dato osservato. Se l'origine non viene registrata, l'incertezza aggiuntiva scompare dal calcolo e l'intervallo finale risulta più stretto di quanto dovrebbe. Ogni valore stimato dovrebbe perciò conservare l'indicazione del metodo con cui è stato ottenuto e, dove disponibile, una misura della sua affidabilità.

Manipolazione dei dati di ingresso

I modelli che suppliscono ai dati mancanti dipendono da fonti esterne che un attaccante può tentare di alterare. La tassonomia del NIST sull'adversarial machine learning descrive, per i sistemi predittivi, attacchi di evasion, di poisoning e alla privacy, e per quelli generativi il misuse e la prompt injection [39]. Il report IBM e Ponemon indica che il 92% delle organizzazioni che hanno subito un breach legato a sistemi di AI non disponeva di adeguati controlli di accesso su tali sistemi [10].

Qualità delle prove

Le prove più solide riguardano l'accuratezza dei singoli modelli su dati pubblici. Non ho trovato studi indipendenti che misurino se le stime quantitative prodotte con queste tecniche siano più vicine alle perdite poi osservate rispetto a quelle prodotte senza. La mia verifica non colma la lacuna, perché riguarda singoli componenti in un dominio vicino. Finché il confronto con le perdite osservate non sarà disponibile, il beneficio per l'analisi quantitativa può dirsi dimostrato per i singoli componenti e atteso, ma non misurato, per il risultato complessivo.

Condizioni d'impiego

Dai limiti descritti e dalla verifica derivano alcune condizioni. La stima a priori ricavata da fonti esterne va dichiarata e motivata insieme al contenuto informativo che le si attribuisce, perché con pochi dati locali è questo parametro a determinare il risultato: dove esistono dati di più unità va stimato, dove non esistono va sottoposto ad analisi di sensibilità. Le osservazioni più vecchie vanno pesate meno di quelle recenti. I valori stimati vanno tenuti distinti da quelli osservati lungo tutta la catena di calcolo e trattati come distribuzioni. L'accuratezza dei modelli va verificata su un campione dei dati dell'organizzazione e ricontrollata nel tempo. Per le grandezze a coda pesante e soggette a mancanza non casuale la media va sostituita da quantili e scenari, accompagnati da un'analisi di sensibilità. Gli organi di gestione approvano le misure e ne sovrintendono l'attuazione, come prevede la NIS2 [1]; perché possano farlo con cognizione di causa, un analista deve essere in grado di ricostruire l'origine di ciascuna stima.

11 Conclusioni

L'analisi quantitativa del rischio cyber è limitata soprattutto dalla disponibilità dei dati, e le tecniche di AI incidono su questo limite in modo selettivo. Dove l'informazione esiste in forma aggregata o non strutturata, la rendono utilizzabile: la probabilità di exploitation stimata da EPSS, gli attributi delle vulnerabilità ricavati dal testo e gli eventi estratti dai rapporti sono dati che la singola organizzazione non potrebbe produrre. Dove l'informazione locale è scarsa, la combinazione con stime esterne riduce l'incertezza in misura calcolabile. Nell'esempio che ho proposto l'ampiezza dell'intervallo sulla frequenza diminuisce del 40%, e una decisione di investimento che i soli dati locali lasciavano incerta può essere assunta con una probabilità di convenienza di quasi il 90%. La verifica su dati pubblici conferma il meccanismo e ne precisa le condizioni: con 50 osservazioni l'errore di stima si dimezza, ma il risultato dell'esempio dipende dal contenuto informativo attribuito alla stima esterna, e con quello appreso dai dati la probabilità di convenienza scende al 52%.

Il contributo si arresta dove il dato manca per ragioni legate al suo valore. Le perdite non dichiarate e gli eventi estremi restano fuori dalla portata delle procedure di completamento, e i dati sintetici non vi pongono rimedio. Per queste grandezze l'analisi quantitativa richiede ipotesi dichiarate e verifiche di sensibilità, che l'impiego dell'AI non rende superflue. Sulle severità reali esaminate le code sono così pesanti che la perdita media non è stimabile in modo stabile, e l'analisi deve poggiare su quantili e scenari.

Nella pratica il beneficio dipende dalle scelte di metodo già indicate: dichiarare le stime a priori e il peso che si attribuisce loro, pesare meno le osservazioni più vecchie, tenere distinti i valori stimati da quelli osservati, verificare i modelli sui propri dati e ripetere la verifica nel tempo. Per la ricerca resta da misurare l'effetto complessivo, confrontando nel tempo le stime ottenute con e senza queste tecniche con le perdite effettivamente registrate. La decisione del NIST di limitare l'arricchimento delle CVE rende la questione più urgente, perché lascia a soggetti diversi dall'ente, e ai modelli che questi impiegano, una parte del lavoro da cui dipendono le stime di vulnerabilità.

Riferimenti

Ho consultato le risorse online il 21 e il 22 novembre 2026. La numerazione segue l'ordine di prima citazione.

1. Direttiva (UE) 2022/2555 del Parlamento europeo e del Consiglio, del 14 dicembre 2022 (NIS 2), GU L 333 del 27.12.2022. EUR-Lex. Testo degli artt. 21 e 23 consultato nella riproduzione di Advisera.
2. Regolamento (UE) 2022/2554 del Parlamento europeo e del Consiglio, del 14 dicembre 2022 (DORA). Sintesi EUR-Lex.
3. Cremer, F., Sheehan, B., Fortmann, M., Kia, A. N., Mullins, M., Murphy, F., Materne, S. (2022). Cyber risk and cybersecurity: a systematic review of data availability. *The Geneva Papers on Risk and Insurance: Issues and Practice*, 47(3), 698-736. DOI 10.1057/s41288-022-00266-6. Scheda University of Limerick.
4. Clusit (2026). *Rapporto Clusit, I semestre 2026: record di incidenti cyber nel mondo*. Comunicato stampa, Security Summit, Verona, 28 ottobre 2026. Testo del comunicato.
5. NIST (2026). *NIST Updates NVD Operations to Address Record CVE Growth*, 15 aprile 2026, aggiornato il 17 aprile. nist.gov.
6. CVE Program (2026). *Metrics: Published CVE Records*. cve.org. Valori annuali letti nelle tabelle di Invadel e ScanTitan, che citano la stessa pagina.
7. Verizon Business (2026). *2026 Data Breach Investigations Report*, 19ª edizione. Executive Summary; comunicato del 19 maggio 2026.
8. ENISA (2026). *ENISA Threat Landscape 2026*. Comunicato del 22 settembre 2026.
9. Wang, J., Neil, M., Fenton, N. (2020). A Bayesian network approach for cybersecurity risk assessment implementing and extending the FAIR model. *Computers & Security*, 89, 101659. DOI 10.1016/j.cose.2019.101659. Scheda Queen Mary University of London.
10. IBM e Ponemon Institute (2026). *Cost of a Data Breach Report 2026*. Pagina del report; comunicato del 29 luglio 2026; analisi di L. Kessem, IBM X-Force.
11. FIRST. *The EPSS Model*. first.org.
12. ISO/IEC 27005:2022. *Information security, cybersecurity and privacy protection. Guidance on managing information security risks*, 4ª edizione. iso.org.
13. NIST (2024). *The NIST Cybersecurity Framework (CSF) 2.0*, NIST CSWP 29, 26 febbraio 2024. nvlpubs.nist.gov.
14. Cox, L. A. (2008). What's Wrong with Risk Matrices? *Risk Analysis*, 28(2), 497-512. DOI 10.1111/j.1539-6924.2008.01030.x. Scheda RePEc.
15. Woods, D. W., Böhme, R. (2021). SoK: Quantifying Cyber Risk. *2021 IEEE Symposium on Security and Privacy*, 211-228. Versione ad accesso aperto.

16. The Open Group (2021). *Risk Analysis (O-RA) Standard, Version 2.0.1* e *Risk Taxonomy (O-RT) Standard, Version 3.0.1*, Open FAIR Body of Knowledge. opengroup.org.

17. Megas, K., Cuthill, B., Dotter, M., et al. (2025). *Cybersecurity Framework Profile for Artificial Intelligence (Cyber AI Profile)*, NIST IR 8596, Initial Preliminary Draft, 16 dicembre 2025. csrc.nist.gov.

18. Eling, M., Wirfs, J. (2019). What are the actual costs of cyber risk events? *European Journal of Operational Research*, 272(3), 1109-1119. DOI 10.1016/j.ejor.2018.07.021. Scheda RePEc.

19. Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592. Scheda Harvard DASH.

20. Jacobs, J., Romanosky, S., Edwards, B., Adjerid, I., Roytman, M. (2021). Exploit Prediction Scoring System (EPSS). *Digital Threats: Research and Practice*. DOI 10.1145/3436242. Preprint arXiv:1908.04856.

21. Shahid, M. R., Debar, H. (2021). CVSS-BERT: Explainable Natural Language Processing to Determine the Severity of a Computer Security Vulnerability from its Description. *20th IEEE International Conference on Machine Learning and Applications* (ICMLA). arXiv:2111.08510.

22. Alam, M. T., Bhusal, D., Nguyen, L., Rastogi, N. (2024). CTIBench: A Benchmark for Evaluating LLMs in Cyber Threat Intelligence. *Advances in Neural Information Processing Systems 37* (NeurIPS 2024), *Datasets and Benchmarks Track*. proceedings.neurips.cc; arXiv:2406.07599.

23. Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S. B., Schirrmeister, R. T., Hutter, F. (2025). Accurate predictions on small data with a tabular foundation model. *Nature*, 637(8045), 319-326. DOI 10.1038/s41586-024-08328-6. PubMed Central.

24. Angelelli, M., Arima, S., Catalano, C., Ciavolino, E. (2024). A robust statistical framework for cyber-vulnerability prioritisation under partial information in threat intelligence. *Expert Systems with Applications*, 255, 124572. DOI 10.1016/j.eswa.2024.124572. arXiv:2302.08348.

25. Jordon, J., Szpruch, L., Houssiau, F., Bottarelli, M., Cherubin, G., Maple, C., Cohen, S. N., Weller, A. (2022). *Synthetic Data - what, why and how?* Rapporto commissionato dalla Royal Society. arXiv:2205.03257.

26. fkie-cad (2026). *nvd-json-data-feeds: Community reconstruction of the deprecated* JSON NVD *Data Feeds*, tag v2026.11.21-010015, commit 0187fe96. Repository GitHub.

27. CISA (2026). *Known Exploited Vulnerabilities Catalog*, versione 2026.11.20. Repository GitHub.

28. ARP Syndicate (2025). *cve-scores: EPSS & VEDAS Score Aggregator for CVEs*, istantanee del 10 aprile e del 13 luglio 2025. Repository GitHub.

29. Verizon RISK Team (2026). *VERIS Community Database* (VCDB), stato al 20 novembre 2026. Repository GitHub.

30. FIRST. *Common Vulnerability Scoring System v3.1: Specification Document*. first.org.

31. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. jmlr.org.

32. Maillart, T., Sornette, D. (2008). Heavy-Tailed Distribution of Cyber-Risks. arXiv:0803.2256.

33. Decreto legislativo 4 settembre 2024, n. 138; determinazioni ACN n. 164179/2025, n. 379907/2025 e n. 127434/2026. Sintesi di CMS.

34. Regolamento (UE) 2024/1689 (AI Act), modificato dal regolamento (UE) 2026/1744 dell'8 luglio 2026. Sintesi EUR-Lex; regolamento modificativo; art. 15.

35. NIST (2023). *Artificial Intelligence Risk Management Framework* (AI RMF 1.0), NIST AI 100-1. nvlpubs.nist.gov.

36. ISO/IEC 42001:2023. *Information technology. Artificial intelligence. Management system*. iso.org.

37. Arp, D., Quiring, E., Pendlebury, F., Warnecke, A., Pierazzi, F., Wressnegger, C., Cavallaro, L., Rieck, K. (2022). Dos and Don'ts of Machine Learning in Computer Security. *31st USENIX Security Symposium*, 3971-3988. usenix.org.

38. Sommer, R., Paxson, V. (2010). Outside the Closed World: On Using Machine Learning for Network Intrusion Detection. *2010 IEEE Symposium on Security and Privacy*, 305-316. IEEE Xplore.

39. Vassilev, A., Oprea, A., Fordyce, A., Anderson, H., Davies, X., Hamin, M. (2025). *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations*, NIST AI 100-2e2025. csrc.nist.gov.

L'autore



Vincenzo Calabrò
Funzionario alla Sicurezza CIS · Ministero dell'Interno
Digital Forensics Analyst

È laureato in Ingegneria Informatica e in Sicurezza Informatica presso le Università di Roma La Sapienza e di Milano. Ha indirizzato la propria formazione nei settori della cyber security e della digital forensics, conseguendo i diplomi di perfezionamento in Data Protection e Data Governance, in Criminalità Informatica e Investigazioni Digitali e in Big Data e Artificial Intelligence. Ha inoltre ottenuto l'Advanced Cybersecurity Graduate Certificate alla School of Engineering della Stanford University e i Professional Certificate in Information Security; Incident Response Process; Digital Forensics; Cybersecurity Engineering and Software Assurance; Applied Data Science for Cybersecurity; Artificial Intelligence (AI) for Cybersecurity al Software Engineering Institute della Carnegie Mellon University.

Dal 1992 è nei ruoli del Ministero dell'Interno, ove ricopre l'incarico di Funzionario alla Sicurezza CIS: in tale veste contribuisce alla valutazione dei rischi cyber, all'implementazione delle misure di sicurezza e alla risoluzione di incidenti informatici. Offre altresì consulenza tecnica nel campo della digital forensics per l'Autorità giudiziaria, la Polizia giudiziaria e gli studi legali.